

A. Y. KHINCHIN

A
Course
of
MATHEMATICAL
ANALYSIS

A COURSE OF MATHEMATICAL ANALYSIS

In the Series

INTERNATIONAL MONOGRAPHS ON
ADVANCED MATHEMATICS & PHYSICS

FURTHER TITLES IN THIS SERIES

1. **Mikhlin:** Linear Integral Equations.
2. **Korovkin:** Linear Operators and the Theory of Approximation.
3. **Khinchin:** Analytical Theory of Statistical Physics.
4. **El'sgolt's:** Differential Equations.
5. **Krasnoselsky:** Convex Functions and Orlich's Spaces.
6. **Lavrentyev-Shabat:** Methods of the Theory of Functions of a Complex Variable.
7. **Lusternik & Sobolev:** Elements of Functional Analysis.
8. **Blokhin:** Methods of X-Ray Spectrum Analysis.
9. **Chechulin:** Wave Processes, Optics and Elements of Atomic and Nuclear Physics.
10. **Romanovski:** Fourier Series, Field Theory and Laplace Transform.
11. **Fuchs & Shabat:** Functions of a Complex Variable and some of their applications.
12. **Goldanskii:** Counting Statistics in Recording Nuclear Particles.
13. **Shtokalo:** Linear Differential Equations with Variable Coefficients.
14. **Gradshtein:** Direct and Reverse Theorems.
15. **Novozhilov:** Elementary Particles.

A COURSE OF MATHEMATICAL ANALYSIS

by
Academician A. KHINCHIN
Moscow University
U. S. S. R.

Translated from the third Russian Edition (1957)

1960

Hindustan Publishing Corp. (India)
DELHI

Published by :

HINDUSTAN PUBLISHING CORP. (INDIA)

6, U. B. Jawahar Nagar, DELHI-6.

Copyright

©

1960

Copyright reserved by the Hindustan Publishing Corp. (India).

This book, or parts thereof, may not be reproduced in any form without the written permission of the publishers.

PRINTED IN INDIA AT THE CENTRAL ELECTRIC
PRESS, KAMLA NAGAR, DELHI-6. PHONE : 20123

PREFACE TO THE FIRST RUSSIAN EDITION

This course of mathematical analysis is a text-book for students of mechanico-mathematical and physico-mathematical faculties of our universities (and to some extent of pedagogical institutes as well) ; it is intended as the main text-book in the study of a science which appears in the curriculum under the heading of mathematical analysis and which deals with the theory of limits, infinite series and differential calculus with simple applications of these subjects. The necessity for such a text-book arose as most of the text-books on mathematical analysis published in this country have not fully satisfied the above requirements. Text-books which by their briefness and simplicity of treatment are within the reach of the average student are usually either obsolete or of lower scientific level than is required for the training of specialized mathematicians ; other text-books which keep on the modern level are usually very bulky and their contents reach far beyond the scope of the current curriculum so that the average first and second years students are unable to benefit by them. It was therefore necessary to write a text-book whose contents would only include the strict requirements of the current curriculum and which would, at the same time, fully conform to the modern scientific standard.

In attempting to make this text-book as brief as possible, I have selected the minimum necessary material and avoided all slackness in treatment. On the other hand, to help the student as far as possible, only the minimum detail is given throughout this course. I have not used words sparingly while trying to explain the line of argumentation. The relationship among various concepts, theorems, problems and theories, their importance and method of application in the applied fields and industry, as well as many other points of mathematical analysis are, in many cases dealt here more comprehensively and systematically than is usually done in other more extensive text-books. I have tried to make the student ready to appreciate the introduction of new concepts and construction of new theories and make him accept them naturally and inevitably. I think that it has been only thus possible to maintain the continuity of interest of the student and make him absorb the subject in an informal manner.

An experienced reader will probably find that the theory of limits has been discussed in its full detail in chapters II, III and IV.

This theory is traditionally presented to secondary school students on the XVIII century level ; university text-books mathematical analysis immediately give the modern treatment of the theory of limits with all the ϵ 's and δ 's, and this is often preceded by a chapter devoted to the general theory of real numbers—a subject which does not, in fact, belong to analysis but to the theory of numbers and the theory of sets. As a result, the student thinks that the new “university” treatment of limiting processes has nothing in common with those limits which he has known at school. In the second place—and this is even more important—this method can rob the student of elements of mathematical analysis as a live, dynamic and dialectic science which find their place in the history of scientific development and which have even today many of their practical applications. These undesirable effects which I had occasion to observe in many instances during my career as a teacher prompted me to use in this text a completely new system for treating the theory of limits. This system essentially involves the following. At first (chapter II) the theory of limits is mainly based on an elementary but not a completely formal basis, and concepts like “process” or “moments” which are not fully defined anywhere are systematically used. Only afterwards the necessity of formalisation has been emphasised, and the fundamental mathematical types of processes defined (chapter III). Then the attention of the student is drawn to the necessity of constructing a general theory of real numbers, and such a theory is, in fact, given in chapter IV. This method of treatment, which I had occasion to test three times in practice has the useful advantage that it creates in the mind of the student a gradual transition from the “school” theory of limits to its “university” treatment, and all stages of this treatment, are fully explained. At the same time it enables him to create at the beginning and maintain throughout the course the basic concepts of mathematical analysis as a live and dynamic subject and concentrate on the formally logical refinements of this subject, which is its due.

So far as the general theory of real numbers is concerned, I have found it necessary to convince the reader of its significance and quote one of the possible principles explaining the existence of imaginary numbers (the limit of a monotone bounded sequence). Only then I have enumerated the basic problems which the theory has to deal (order in a continuum, definitions and rules of algebraic methods) ; at this point I have also given few examples of their solution, indicating briefly that the theory of numbers can be applied quite satisfactorily to these problems and that, in future, we shall

deliberately use the results provided by this theory. The future mathematicians will be able to learn in other more detailed courses the fact that the theory of numbers can solve all these problems; this problem is hardly of any interest to the future mechanic, physicist, or astronomer. In any case, I do not consider it possible to attract the attention of a varied audience, either in my lectures or in this book, to the study of a large chapter the contents of which have no immediate connection with mathematical analysis.

The further treatment of the subject follows, in its main outlines, certain well defined methods. I am sorry to say that in editing the last three chapters (multiple, curvilinear and surface integrals) my attempts to make the treatment absolutely formal and, at the same time, easily accessible did not meet with success, as far as I am able to judge. I could not avoid compromises by sacrificing either the formality or the briefness and accessibility of the arguments. If this course is received with favour, then it will undoubtedly be necessary to work further on these chapters in future editions.

A few problems given in this course are valuable only as illustrations, but they are not intended as a method of instruction. The number and character of these problems correspond to what a lecturer can convey during his lectures. I had no intention of including the material for practical (group) lessons in this course of analysis. Obviously, anyone studying this book should simultaneously use a good book for problems. For this purpose the recently published "Problem Book on Mathematical Analysis" by B.P. Demidovich (Gostekhizdat, 1952) is particularly suitable. For the convenience of certain classes of readers I have indicated in many paragraphs a few problems appearing in the above book which I especially recommend. I must, however, warn the reader that these problems are, in a majority of cases, insufficient for acquiring the necessary skill; a further choice of examples should be left to the teacher-in-charge of practical classes.

A competent reader will readily note that the order in which individual subjects are treated in this book is in no way compulsory and can be easily altered in many instances; for example 1) some geometrical applications of differential calculus (chapter XXIII) can be given (and are, in fact, usually given) much earlier, and 2) the integral test for convergence of series must not be postponed until the theory of generalised integrals is dealt with (chapter XXV); but it can be given with the treatment of series of constant signs (chapter XVIII, § 68).

It is my pleasant duty to express my sincere and deep gratitude to my colleagues of the Faculty of Mathematical Analysis at the Moscow, Leningrad and Kiev Universities for their valuable help given by reading the manuscript (or its individual chapters) and for their remarks and suggestions which have mostly led to significant improvement in the treatment of the subject. In this respect I am particularly grateful to Prof. L.A. Tamarin (Moscow) and Prof. G.E. Shilov (Kiev). Finally, I want to thank the editor of my book, O.N. Golovin, for his competent and considerable work devoted to this book; his many valuable suggestions have helped considerably to improve its contents.

Moscow.

24 February, 1953.

A. KHINCHIN

PREFACE TO THE SECOND RUSSIAN EDITION

The second edition of this book is mainly printed from blocks, and corrections of many individual mistakes and errors have been done by the author; in some cases attempts have been made to improve the treatment of the subject. In this respect I have been greatly helped by a detailed criticism of this book sent to me by the Faculty of Mathematical Analysis at the Rostov University (under the chairmanship of Prof. F.D. Gakhov); I am deeply grateful to all the members of this Faculty. I am also very thankful to Academician A.N. Kolmogorov and Prof. A.D. Myshkis (Minsk) for pointing out some mistakes.

The "Problem Book on Mathematical Analysis" by B.P. Demidovich which has been frequently referred in this book has appeared in its second edition in 1954 with a fundamental revision of the numbering of problems. In the present edition of this course the numbering of all recommended exercises refer to the first edition of this "Problem Book."

*Moscow,
19 December, 1954.*

A. KHINCHIN

CONTENTS

Chapter 1. FUNCTIONS	...	1
§ 1. Variables		1
§ 2. Functions	...	3
§ 3. The region of definition of a function	...	6
§ 4. Functions and formulae	...	7
§ 5. The geometrical representation of functions	...	11
§ 6. Elementary functions	...	13
Chapter 2. ELEMENTARY THEORY OF LIMITS	...	18
§ 7. Infinitesimal quantities	...	18
§ 8. Operations with infinitesimal quantities	...	23
§ 9. Infinitely large quantities	...	26
§ 10. Quantities which tend to limits	...	29
§ 11. Operations with quantities which tend to limits	...	33
§ 12. Infinitesimal and infinitely large quantities of different orders.	...	39
Chapter 3. THE DEVELOPMENT OF THE ACCURATE THEORY OF LIMIT TRANSITION	...	45
§ 13. The mathematical definition of a process	...	45
§ 14. The accurate concept of limits	...	47
§ 15. The development of the concept of limit transitions	...	52
Chapter 4. REAL NUMBERS	...	56
§ 16. Necessity of producing a general theory of real numbers	...	56
§ 17. Construction of a continuum	...	59
§ 18. Fundamental lemmas	...	69
§ 19. Final points in connection with the theory of limits	...	74
Chapter 5. CONTINUOUS FUNCTIONS	...	79
§ 20. Definition of continuity	...	79
§ 21. Operations with continuous functions	...	84
§ 22. Continuity of a composite function	...	85
§ 23. Fundamental properties of continuous functions	...	87
§ 24. Continuity of elementary functions	...	94

Chapter 6. DERIVATIVES	...	98
§ 25. Uniform and non-uniform variation of functions	...	98
§ 26. Instantaneous velocity of non-uniform movement	...	101
§ 27. Local density of a heterogeneous rod	...	106
§ 28. Definition of a derivative	...	108
§ 29. Laws of differentiation	...	110
§ 30. The existence of functions and their geometrical illustration	...	123
Chapter 7. DIFFERENTIALS	...	128
§ 31. Definition and relationship with derivatives	...	128
§ 32. Geometrical illustration and laws for evaluation	...	132
§ 33. Invariant character of the relationship between a derivative and a differential	...	134
Chapter 8. DERIVATIVES AND DIFFERENTIALS OF HIGHER ORDERS	...	136
§ 34. Derivatives of higher orders	...	136
§ 35. Differentials of higher orders and their relationship with derivatives	...	139
Chapter 9. MEAN VALUE THEOREMS	...	142
§ 36. Theorem on finite increments	...	142
§ 37. Evaluation of limits of ratios of infinitely small and infinitely large quantities	...	147
§ 38. Taylor's formula	...	154
§ 39. The last term in Taylor's formula	...	158
Chapter 10. APPLICATION OF DIFFERENTIAL CALCULUS TO ANALYSIS OF FUNCTIONS	...	164
§ 40. Increasing and decreasing of functions	...	164
§ 41. Extrema	...	167
Chapter 11. INVERSE OF DIFFERENTIATION	...	175
§ 42. Concept of primitives	...	175
§ 43. Simple general methods of integration	...	182
Chapter 12. INTEGRAL	...	193
§ 44. Area of a curvilinear trapezium	...	193
§ 45. Work of a variable force	...	198
§ 46. General concept of an integral	...	201
§ 47. Upper and lower sums	...	204
§ 48. Integreability of functions	...	207

Chapter 13. RELATIONSHIP BETWEEN AN INTEGRAL AND A PRIMITIVE	...	213
§ 49. Simple properties of integrals	...	213
§ 50. Relationship between an integral and a primitive	...	218
§ 51. Further properties of integrals	...	223
Chapter 14. THE GEOMETRICAL AND MECHANICAL APPLICATIONS OF INTEGRALS	...	230
§ 52. Length of an arc of a plane curve	...	230
§ 53. Lengths of arcs of curves in space	...	241
§ 54. Mass, centre of gravity and moments of inertia of a material plane curve	...	242
§ 55. Capacities of geometrical bodies	...	247
Chapter 15. APPROXIMATE EVALUATION OF INTEGRALS	...	254
§ 56. Problematic set up	...	254
§ 57. Method of trapeziums	...	257
§ 58. Method of parabolas	...	262
Chapter 16. INTEGRATION OF RATIONAL FUNCTIONS	...	265
§ 59. Algebraical introduction	...	265
§ 60. Integration of simple fractions	...	274
§ 61. Ostrogradskij's method	...	277
Chapter 17. INTEGRATION OF THE SIMPLE RATIONAL AND TRANSCENDENTAL FUNCTIONS	...	282
§ 62. Integration of functions of the type		
$R \left(x, \sqrt[n]{\frac{ax+b}{cx+d}} \right)$...	282
§ 63. Integration of functions of the type		
$R \left(x, \sqrt{ax^2+bx+c} \right)$...	284
§ 64. Primitives of binomial differentials	...	287
§ 65. Integration of trigonometrical differentials	...	289
§ 66. Integration of differentials containing exponential functions	...	294
Chapter 18. NUMERICAL INFINITE SERIES	...	297
§ 67. Fundamental concepts	...	297
§ 68. Series with constant signs	...	305
§ 69. Series with variable signs	...	316
§ 70. Operations with series	...	320
§ 71. Infinite products	...	326

Chapter 19. INFINITE SERIES OF FUNCTIONS	...	333
§ 72. Region of convergence of a series of functions	...	333
§ 73. Uniform convergence	...	335
§ 74. The continuity of the sum of a functional series	...	340
§ 75. Term-by-term integration and differentiation of series	...	344
Chapter 20. POWER SERIES AND SERIES OF POLYNOMIALS	...	351
§ 76. Region of convergence of a power series	...	351
§ 77. Uniform convergence and its consequences	...	357
§ 78. Expansion of functions into power series	...	361
§ 79. Series of polynomials	...	369
§ 80. Theorem of Weierstrass	...	372
Chapter 21. TRIGONOMETRICAL SERIES	...	377
§ 81. Fourier coefficients	...	377
§ 82. Average approximation	...	383
§ 83. Dirichlet-Liapunov theorem on closed trigonometrical systems	...	388
§ 84. Convergence of Fourier series	...	394
§ 85. Generalised trigonometrical series	...	396
Chapter 22. DIFFERENTIATION OF FUNCTIONS OF SEVERAL VARIABLES	...	400
§ 86. Continuity of functions of several independent variables	...	400
§ 87. Two-dimensional continuum	...	403
§ 88. Properties of continuous functions	...	408
§ 89. Partial derivatives	...	410
§ 90. Differentials	...	413
§ 91. Derivatives in arbitrary directions	...	419
§ 92. Differentiation of composite and implicit functions	...	422
§ 93. Homogeneous functions and Euler theorem	...	427
§ 94. Partial derivatives of higher orders	...	429
§ 95. Taylor's formula for functions of two variables	...	433
§ 96. Extrema	...	438
Chapter 23. SOME SIMPLE GEOMETRICAL APPLICATIONS OF DIFFERENTIAL CALCULUS	...	443
§ 97. Equations of tangent and normal to a plane curve	...	443
§ 98. Tangential line and normal plane to a curve in space	...	446

§ 99.	Tangential and normal planes to a surface	...	448
§ 100.	Direction of convexity and concavity of a curve	...	451
§ 101.	Curvature of a plane curve	...	453
§ 102.	Tangential circle	...	458
Chapter 24.	IMPLICIT FUNCTIONS	...	462
§ 103.	The simplest problem	...	462
§ 104.	The general problem	...	469
§ 105.	Ostrogradskij's determinant	...	475
§ 106.	Conditional extremum	...	483
Chapter 25.	GENERALISED INTEGRALS	...	491
§ 107.	Integrals with infinite limits		491
§ 108.	Integrals of unbounded functions	...	504
Chapter 26.	INTEGRALS OF PARAMETRIC FUNCTIONS	...	514
§ 109.	Integrals with finite limits	...	514
§ 110.	Integrals with infinite limits	...	526
§ 111.	Examples	...	535
§ 112.	Euler's integrals	...	541
§ 113.	Stirling's formula	...	548
Chapter 27.	DOUBLE AND TRIPLE INTEGRALS	...	557
§ 114.	Measurable plane figures	...	557
§ 115.	Volumes of cylindrical bodies	...	567
§ 116.	Double integral	...	571
§ 117.	Evaluation of double integrals by means of two simple integrations	...	576
§ 118.	Substitution of variables in double integrals	...	584
§ 119.	Triple integrals	...	590
§ 120.	Applications	...	593
Chapter 28.	CURVILINEAR INTEGRALS	...	602
§ 121.	Definition of a plane curvilinear integral	...	602
§ 122.	Work of a plane field of force	...	610
§ 123.	Green's formula	...	612
§ 124.	Application to differentials of functions of two variables	...	617
§ 125.	Curvilinear integrals in space	...	622
Chapter 29.	SURFACE INTEGRALS		626
§ 126.	The simplest case	...	626
§ 127.	General definition of surface integrals	...	630

§ 128. Ostrogradskij's formula	...	637
§ 129. Stoke's formula	...	642
§ 130. Elements of the field theory	...	647
CONCLUSION—Short historical sketch	...	653
[INDEX	...	665

CHAPTER I

FUNCTIONS

§ 1. Variables

The introduction of the *variable* was a decisive step in mathematics. Thus *movement* and *dialectics* were introduced in mathematics. (*F. Engels, Dialectics of Nature, Gospolizdat, 1948, p. 208.*)

Elementary mathematics—the mathematics of constants—revolves, as it were, within limits of formal logics; the mathematics of variables, which is chiefly concerned with infinitely small quantities, essentially involves the application of dialectics to mathematical relationships. (*F. Engels, Anti-During, Gospolizdat,, 1948, p. 127.*)

When we observe a natural phenomenon or the course of a technical process we can usually note the different behaviour of quantities involved in this phenomenon or process. Some quantities do not change in the course of the process, *i.e.* they remain “constant”, while others are subjected to greater or lesser change—they become greater or smaller—*i.e.* they are “variable”. If we heat a gas confined in a closed vessel its volume remains constant; the number of molecules of the gas also remains constant; on the other hand the temperature of the gas, and its pressure will grow and become increasingly greater. The picture becomes even more varied if instead of considering this simple laboratory experiment we consider a complicated technical process. Let us consider, for example, the flight of an aeroplane. Many different quantities are involved in this phenomenon. Some of these remain constant throughout the flight; *e.g.* the number of passengers, the weight of their luggage, the span of the wings of the aeroplane, and many others. However, this process also involves many other quantities which alter during the process by becoming greater or smaller. Such are, for example, the distance of the aeroplane from the point of departure and from its destination, its height above the earth, the supply of fuel, the temperature, pressure and humidity

of the surrounding air, and many others. The above summary shows that these variable quantities are most important in economical and technical calculations connected with this process. This can readily be understood. Nature involves continuous changes and the practical life of man is directed towards changing his surroundings. For this reason processes in which nothing, or almost nothing, changes have little to offer scientifically and are of no practical interest. According to the dialectic principles of nature study, we should study not so much the instantaneous aspect of phenomena but their changes in time; from the dialectic point of view we are not so much interested in the given aspect of a phenomenon but in the general course of the phenomenon, *i.e.* we are interested how and what changes if this phenomenon took place from time to time. Mathematics, in as far as it is a real tool in nature study, should be able to provide an apparatus which would enable one to study systematically any changes in quantities which take place in nature and in technical processes.

Mathematical analysis is such an apparatus and, in the widest sense of the word, can be called the mathematical science of variables.

Hence the first basic concept in mathematical analysis is the variable quantity or, as it is usually said in mathematics, the concept of the *variable*. By this we mean quantities which acquire varying values, either greater or smaller, in the course of the given process; at different stages of a given process the values of this quantity are, generally speaking, different. Without going into further details we know from everyday experience that the character and manner in which quantities change can follow a very diverse course; some quantities increase continuously; other quantities, on the other hand, decrease continuously; still others change in a vibrating manner by first growing and then diminishing (the distance of the Earth from the Sun, the deflection of a pendulum from the vertical position); if we assume that the given quantity grows continuously, it can do so either very rapidly or very slowly, *i.e.*, the pace of its growth can become quicker or slower. Mathematical analysis in its widest sense enables us to study systematically these and other characteristic changes of quantities in our surroundings; it introduces a definite pattern into the enormous number of various types of changes and finds common laws which govern changes of various types.

In mathematics every quantity involved in a phenomenon, irrespective of whether it is a constant or a variable, is usually denoted by a single letter. Thus, for example, if a quantity is denoted by the letter x or by the letter a , then this fact by itself gives no indication as

to whether this quantity is a constant or a variable ; therefore the way in which this quantity changes must be stressed separately. Furthermore it is very important to keep in mind the fact that without the knowledge of the process (*phenomenon*) in hand, we cannot, generally speaking, know whether this or another quantity is a constant or a variable. The same quantity can be a constant in one process and a variable in another process ; thus, for example, if we rotate a circle of radius r about a straight line without changing its radius (first process) then the area of this circle πr^2 will be constant ; if, however, we keep the centre of the circle stationary and increase its radius (second process), then the area of the circle will grow, *i.e.* it will be a variable.

In mathematical analysis the well-known geometrical representation of numbers by points on a straight line (the so-called “number line”) is widely used. If we denote the origin by O and a unit of length on the straight line, then we can represent an arbitrary number α by a point at a distance $|\alpha|$ from the point O in a direction which depends on the sign of the number α (generally, if the number line is horizontal, positive numbers are plotted to the right and negative numbers to the left of the point O). Every value of x is a number and can be represented by a point on the number line. If in the given process the value of x is constant then this value is denoted by one and the same point on the number line during the whole process. We can therefore say that a constant is represented by a stationary point on the number line. If, however, the value of x varies during the given process, then its values at different stages of the process are represented by different points on the number line ; in the course of the process the point denoting the value of x changes its position and we can therefore say that a variable is denoted by a *mobile* point on the number line.

§ 2. Functions

Quantities involved in the same phenomenon do not, as a rule, change independently of each other ; usually these quantities are also more or less closely related to one another so that changes in one of these quantities involve corresponding changes in the other quantities. Thus, by increasing the radius of a circle we inevitably also increase its area ; by compressing a gas confined in a vessel (*i.e.* by decreasing the volume occupied by the gas) we also (by keeping the temperature constant) inevitably increase the pressure of the gas ; by adding

*) The symbol $|x|$ denotes the “absolute value of the number x .”

manure to the soil we hope to increase the yield of the harvest, etc. We can see from the above examples that quantities involved in the same phenomenon can bear to one another a more or less close relationship. This relationship is closest in the first example; by knowing the radius r of the circle we can determine its area s uniquely and with absolute accuracy according to the formula $s = \pi r^2$. In the second example the picture is somewhat different; by knowing the volume v occupied by the gas and its absolute temperature T we are able to determine uniquely its pressure p according to the well-known formula :

$$p = \frac{c T}{v},$$

where c is a constant known from physics; however this formula is only accurate with certain (in some cases rather rough) approximations and for more accurate calculations it is necessary to use more complicated formulae which show that when determining the pressure of the gas under real conditions it is insufficient to know its temperature and volume alone, but it is also necessary to take some other quantities into account. This point is even better illustrated in our last example; although it is true to say that the quantity of manure has an undisputed effect on the yield of the harvest it is nevertheless clear that from the knowledge of the quantity of manure used we are unable to forecast the yield of the harvest with any accuracy, for the yield of the harvest, apart from the quantity of manure used, also depends on a series of other factors (for example on meteorological and agrotechnical factors of different kinds).

It is evident that mathematical analysis is mainly concerned with *accurate* relationships existing between quantities, *i.e.* from the knowledge of one group of quantities we are able to determine uniquely and accurately the values of a certain other group of quantities. Consider, for example, the accurate relationship existing in the above formulae

$$s = \pi r^2, \quad p = \frac{c T}{v},$$

where c is a known constant. The value of the radius r of the circle is unique and defines accurately its area s . If we know the quantities T and v then the second of the above formulae enables us to determine quite accurately the corresponding value of p . In the first case the value of s depends *only on one* quantity r ; each value of r corresponds to a definite value of s and every change in the value of r involves a corresponding change in the value of s . The second

example is more complicated ; in order to find the value of p it is not enough to know the value of T or the value of v alone ; the value of p depends on the values of two quantities— T and v ; we must know both values if we are to determine the value of p in accordance with our formula ; to each pair of values v and T corresponds one value of p and changes in the value of p depend on changes in the values of both T and v ; as far as values of v and T are concerned changes in either of them are independent of one another and can take place in any way we like. In the physical sense this means that the given mass of gas can be confined in an arbitrary (within certain limits) volume v and can be heated to an arbitrary (within certain limits) temperature T . But as soon as we have chosen the values of v and T the pressure of the given mass of gas no longer remains arbitrary but is defined uniquely and quite accurately by our formula (we are, of course, omitting the fact that the formula itself requires corrections for real gases).

The above examples are particular cases of the following general scheme. A quantity y involved in a certain process depends on the quantities x_1, x_2, \dots, x_k , which are also involved in the same process ; this dependence is such that to each set of values x_1, x_2, \dots, x_k corresponds a single value of the quantity y ; at the same time the values of x_1, x_2, \dots, x_k are independent of one another, *i.e.* by assuming the values of some of these quantities we can select the values of the remaining quantities quite arbitrarily (usually within certain set limits). This type of dependence of the value of y on the values of x_1, x_2, \dots, x_k is known as *functional dependence* and y is said to be the *function* of x_1, x_2, \dots, x_k ; x_1, x_2, \dots, x_k are, in this case, said to be *independent variables*. Hence in the above examples the value of s is a function of one independent variable r *) and the value of p is a function of two independent variables T and v . To begin with we shall concentrate on the simplest case when $k = 1$, *i.e.* when y is a function of a single independent variable x .

The fact that y is a function of the independent variable x is usually denoted as follows : $y = f(x)$, or $y = \alpha(x)$, or $y = A(x)$, etc. The letter in front of the bracket indicates the functional dependence of y on x and can be selected arbitrarily—the meaning of this notation thereby remains unchanged. Thus the fact that the area of a circle is uniquely determined by its radius can be written down in the form $s = f(r)$, or $s = \alpha(r)$, or $s = A(r)$, etc. Similarly the fact that y is a

*) Frequently instead of using the words “of one” or “of two” it is simply said “one” or “two”, etc.

function of several independent variables x_1, x_2, \dots, x_k can be written in the form of the relationship $y = f(x_1, x_2, \dots, x_k)$, or $y = y(x_1, x_2, \dots, x_k)$, or $y = F(x_1, x_2, \dots, x_k)$, etc. Thus the fact that the pressure p of the given mass of gas is defined uniquely by the values of its volume v and absolute temperature T can be written in the form $p = f(v, T)$, or $p = p(v, T)$, or $p = F(v, T)$, etc. Hence the letter chosen to denote the functional dependence does not tell us anything about the nature of this dependence; the relationship $y = f(x)$ can, in different cases, mean that $y = 3x^2$, or $y = \log x$, or $y = \sin x$, etc. To avoid errors it is only important to see that in the same argument one and the same letter does not symbolise different types of functional dependence. Thus if in a certain process $y = x^2$ and $z = x^3$, then we cannot, of course, write $y = f(x)$ and $z = f(x)$.

On the other hand, in some cases one and the same letter is used to denote a certain quantity and a type of its functional dependence on other quantities [$s = s(r)$ and $y = y(x_1, \dots, x_k)$ in the above examples].

The dialectic method of nature study and the study of technical processes requires that the quantities involved which change during the process should not be studied separately, irrespectively of each other, but should be studied in the same interdependence in which they stand to one another in reality. The mathematical interdependence of real numbers is, in the simpler cases, expressed by the concept of functional dependence. It is therefore clear that if the first basic concept of mathematical analysis is the concept of the variable, as we saw in § 1, then the second concept in the development of the science of variable quantities is the concept of the function. Furthermore the necessity of considering continually variable quantities and their interdependence, both from the scientific and the practical points of view, has made the concept of the function the main object of study in mathematical analysis so that it is quite correct to call this science the general theory of functions.

§ 3. The region of definition of a function

We agreed to call y a function of x if, by assuming the value of x , the value of y is thereby defined uniquely. At the same time it is not necessary that y should be defined for *every* value of x ; the real meaning of the values x and y and the problem in hand determine in every case the values of x which have to be considered. Thus, for example, if y denotes the area of a regular x -sided polygon inscribed in a circle of unit radius, then evidently y is a function of x ; but

from the nature of this problem it is obvious that we are only interested in those values of x which are integers, *e.g.* 3, 4, 5, Similarly $n!$ is a function of n but, by its nature, it becomes devoid of meaning for all numbers other than the integers $n > 0$. The function $y = \log x$ is usually only defined for positive values of x .

If the absolute temperature T of a certain body, given in degrees Centigrade, is the independent variable in a given problem, then, in all probability, we shall not be interested in temperatures below -273 . On the other hand the functions $y = x^2$ or $y = \sin x$, which are given in purely mathematical form, can be defined for all values of x and in practice one meets many such problems which can only be solved if we know how to determine the value of the function for every value of x .

The above examples show clearly that the set of values of the independent variable x for which it is logical and necessary to determine the corresponding values of the function y depends entirely on the nature of the problem in hand. In the choice of this set we are usually guided by mathematical, or sometimes practical, considerations. In any case whenever we deal with an arbitrary function $y = f(x)$ we must keep clearly in mind the set M of those values of the independent variable x for which this function is defined and in cases where there is the slightest doubt it is necessary to mention clearly the appropriate set; for values of x which do not belong to this set the function y is devoid of meaning and is considered to be an undefined function. Therefore the set M is said to be the *region of definition* of the given function.

In view of this it is clear that the set M should be mentioned in the definition of a function :

The quantity y is said to be a function of the quantity x defined by the set M if to every value of x which belongs to the set M there corresponds a definite value of y .

§ 4. Functions and Formulae

Whenever an arbitrary definite function is given in mathematical form it is necessary to define the *relationship* which defines the corresponding value of y for every value of x belonging to the set M . The means of establishing this relationship are, of course, very important from the practical point of view; however, in principle this is only a technical problem of secondary importance. The most convenient method of defining the function $y = f(x)$ is, of course a definition which states clearly the algebraic operations over x and the

order in which they are to be performed so as to obtain the corresponding values of y ; typical examples of this type of definition are simple formulae of the type $y = 3x^2$, $y = \frac{1}{1+x^4}$ etc., with the aid of which it is easy to calculate the values of y for every value of x ; a similar case is provided by the formula

$$n! = 1.2 \dots n,$$

which defines the values of the function $n!$ for all positive integral values of the number n .

However, it is not always possible to define a function in this simple form; and even when this is possible it is not always the most convenient way from a practical point of view. Even such elementary functions as $\log x$, $\sin x$, $\cos x$, etc., are given by formulae which do not give a simple answer to the question of how to find the corresponding value of the function from the given value of x . For example, the function $y = \sin x$ is usually defined by the well-known geometrical representation; the latter convinces us of the existence of a unique, and fully defined function $\sin x$, but it does not give us an immediate method for finding the values of this function. It is therefore necessary to solve this problem by a special method; the fact that the solution of this problem is not simple is obvious from the wide use of *tables* for functions like $\sin x$, $\cos x$, $\log x$, etc.: in these tables the results of calculations of this or other functions for different values of x are given; these results, having once been obtained with considerable effort, are published in the form of tables in order to save scientists and practical workers the unnecessary repetition of calculations.

Below we give a few good examples for defining functions.

Example 1. Let y denote the greatest integer which does not exceed the number x ; it is obvious that the value of y will thus be defined uniquely for every value of x , *i.e.* it is defined as a function of x . This function is usually denoted by the symbol $[x]$ so that, for example

$$[2.5] = 2, \quad [5] = 5, \quad [\pi] = 3, \quad [-\pi] = -4$$

etc. The function $y = [x]$ is of great importance in the theory of numbers and in other branches of mathematics. We can see that it can be defined very simply, but it contains no formulae to indicate the sequence of operations which have to be performed in order to arrive from the given value of x to the corresponding value of $y = [x]$. By the way, it is also possible to express the function $y = [x]$ in terms

of x by means of a “formula”, *i.e.* by means of a series of symbols used in elementary mathematics; however, such a formula would, as a rule, not facilitate in any way the investigation of the function $[x]$ and it is therefore more natural to use its definition without a formula.

The quantity $x - [x]$ is said to be a *fractional part* of the number x and is rather important in the theory of numbers; it is evident that this is a periodical function with a unit period, so that we have

$$0 \leq x - [x] < 1.$$

Example 2. (“Dirichlet’s function”). Assume that $D(x) = 1$ when x is a rational number (*i.e.* an integer or a fraction) and $D(x) = 0$ when x is an irrational number (for example $x = \sqrt{2}$ or $x = \pi$). The function $D(x)$ is defined for all values of x (its region of definition is the whole number line). We can see that its definition is very simple. In order to find the value of $D(x)$ from the given value of x it is only necessary to establish by any arbitrary method whether x is a rational or irrational number; no general method can be given for this purpose—the solution of this problem depends on the way in which x is given; numbers x exist in mathematics which can be accurately defined but which have so far not been defined as rational or irrational numbers; this means that certain values of the function $D(x)$ cannot be evaluated mathematically, but in spite of this the above definition of the function $D(x)$ is quite valid. A “formula” can also be written for the function $D(x)$, *i.e.* it is possible to express it in terms of mathematical symbols in general use. However, such a formula has no practical value since the main properties of Dirichlet’s function can usually be deduced much more easily from the “formulaless” definition given above, whereas with the aid of a formula they cannot be deduced at all or can only be deduced with great difficulty.

The above examples clearly indicate the part played by a formula in analytical expression and in the definition of functional dependence. A formula, when it is simple and convenient for calculations and investigations, can be an invaluable tool in the study and practical application of the given function. However, in cases where a formula cannot be found or where an existing formula is complicated and gives little information, there is no good reason for making a formula the focal point in the study of a function; in many cases “formulaless” investigations are simpler and more productive.

For a long time (during the whole of the XVIII and the begin-

ning of the XIX centuries) the concept of the function was closely linked with a definite analytical expression which, from a useful method in the study of functions, became its exclusive master. This tendency, which is purely formal in character (for the form is the analytical expression and through it the real laws of functional dependence were dictated) was obstinately maintained for many centuries and even today it is not quite obsolete, especially in the applied sciences. A change in outlook, as described in § 3, took place when the definition of the concept of functional dependence became divorced from outside influences; this happened in the middle of the XIX century and is connected with the name of the German mathematician Dirichlet. However, several years before Dirichlet the Russian scientist N. I. Lobachevskij proposed this definition with great clarity.*) In order to distinguish between the formal and the other approach to the definition of the concept of a function and to clarify it still further we give below one more example.

Example 3. Let us assume that

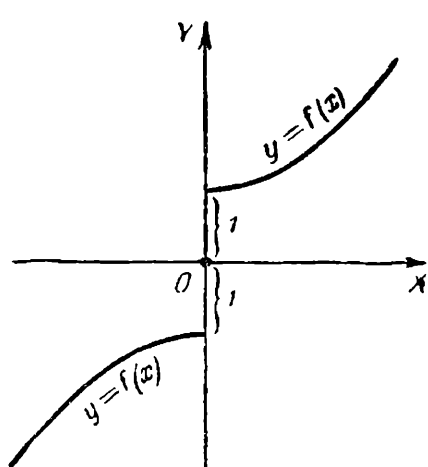


Fig. 1.

$$y = f(x) = \begin{cases} -1 - x^2 & (x < 0), \\ 0 & (x = 0), \\ 1 + x^2 & (x > 0). \end{cases}$$

This means that for negative values of x we must calculate y by the formula $y = -1 - x^2$, and for positive values of x —by the formula $y = 1 + x^2$; when $x = 0$ we have $y = 0$. It follows from our definition that we evidently have here a single function which is defined for all values of x (its region of definition is the whole number line). The graphical representation of this function is given in Fig. 1.

During different stages of the process our function is evaluated with the aid of different formulae; this circumstance has no special significance from the point of view of the definition of our function; it does not alter the fact that irrespective of the value of x there is only one definite corresponding value of y ; this is sufficient to convince us that we are dealing with a single definite function. The

*) Several years earlier the same idea occurred to the Czechoslovakian mathematician Bolzano.

formal point of view, however, connects every function with a definite analytical expression and one would therefore be tempted to say that y is expressed by “different functions” during different stages of the process.

The history of the development of functions and the practical applications of this science have proved without doubt the advantages of the former point of view which has liberated the concept of the function from the burden of a formula as compared to the formal concept which attempts to subject the function to a definite outward form of expression.

Apart from general methodical and logical points of view, this advantage is further based on the fact that functions similar to the function just defined (*i.e.* a function which can be expressed by different formulae during different stages in the process of change of the independent variable) occur quite often in nature and technical problems (particularly in physics, chemistry, thermodynamics, etc.).

§ 5. The geometrical representation of functions

The basic principles in the geometrical representation of functions (the construction of graphs) are studied in secondary schools and we shall here only make a few short remarks in connection with this problem.

By the graph of the given function $f(x)$, we understand a geometric set of points in a plane, the rectangular co-ordinates x and y of which are connected by the relationship $y = f(x)$. If the function $f(x)$ is not unduly complicated then its graph usually represents a more or less straight line in the plane. The fact that each value of x (within the region of definition of the given function) corresponds to a *unique* value of $y = f(x)$ can be illustrated very simply by geometrical means: every straight line which is parallel to the OY-axis intersects the graph of the function $f(x)$ only at one point. Other curves which do not possess this property cannot, generally speaking, serve as the graph of any function of the variable x ; on the other hand, every curve possessing this property is, evidently, the graph of a function, for every unique dependence of y on x does, according to the definition of the concept of a function, represent a certain functional dependence.

The geometrical representation of functions is very important in their study and is therefore a very useful tool in mathematical

analysis and its applications. From the graph of a function we are frequently able to see directly certain characteristics which could otherwise only be revealed by means of lengthy and complicated calculations—by studying the analytical expression or the compiled tables for the given function.

Fig. 2 thus shows that the function $f(x)$ it represents grows (when x increases) along the sections $a_2 a_3$ and $a_4 a_5$ and decreases along the sections $a_1 a_2$ and $a_3 a_4$; to obtain more detailed information, for example, about the way in which the function decreases along the section $a_1 a_2$, we can see directly from the graph that at the beginning (near $x = a_1$) the decrease takes place slowly, but later on it goes rapidly (the steep descent); the function decreases even more rapidly along the section $a_3 a_4$; while the increase of this function is rapid along $a_2 a_3$ and much slower along the section $a_4 a_5$. Within the region under investigation this function reaches its maximum value at the point $x = a_3$ and its minimum value at the point $x = a_4$. We can see clearly from the graph where the function is positive and where it is negative, etc. All this information about the function could only be obtained with much greater difficulty if instead of the graph we had used tables or the analytical expressions of this function.

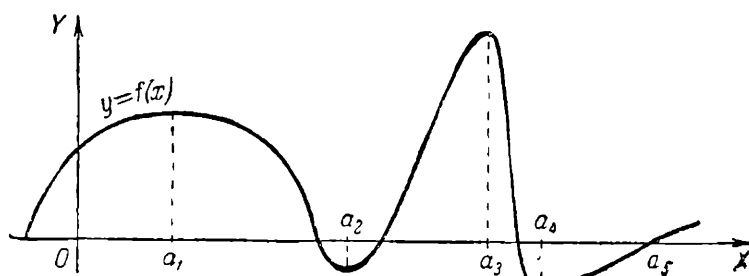


Fig. 2.

The close relationship created by the geometrical representation of the function between the objects of analysis under investigation (functions) and the geometrical objects (curves) enables one not only to use visual representation in the study of the properties of this or other function but, conversely, to use the numerous methods of mathematical analysis for the study of the geometrical properties of this or other curve and, moreover, a series of general geometrical propositions can be established in this way. In future we shall meet many examples of this kind. The connection between analysis and geometry, the first step in which is the principle of geometrical representation of functions, is thus most useful for both mathematical sciences.

§ 6. Elementary functions

In the historical development of the science of functional dependence a small group of functions was isolated from the endless variety of different types; these functions occurred frequently in various problems and thus came to be subjected to detailed study. These functions are the so-called *elementary functions*. Further studies in the development of analysis acquainted scientists with many other more complicated functions which required just as much attention; nevertheless even today elementary functions are the basis in the many applications of analysis; moreover, in the study of other more complicated functional dependencies we do, as a rule, use the well-known properties of this classical group of elementary functions. This group basically coincides with the set of functions which are usually studied in secondary schools; therefore there is no need for us to consider the properties of elementary functions here in greater detail; we shall simply enumerate them and make a few remarks in each case. Some special characteristics of these functions are considered later (§ 24). Elementary functions cannot be isolated as a group by any particular property and, as we have already said above, this small group of functions came to be isolated in the historical development of this science as a natural basis in the study of other more complicated functional dependencies both for the purpose of analysis and for its applications.

1. Polynomials. The simplest form of functional dependence is provided by an arbitrary *polynomial*

$$y = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n,$$

where x is the independent variable, n is an arbitrary natural number and a_0, a_1, \dots, a_n are constants (the "coefficients" of the polynomial). To obtain the values of y for the given value of x we must perform over x and over the constants a series of arithmetical operations (additions, subtractions, multiplications and raising to integral positive powers). Conversely, the result of a series of such operations over x and arbitrary constants can be represented by a polynomial. Therefore polynomials are also known by the name of *rational integral functions*; they are known as integral because the operations we enumerated above do not include division, and rational because these operations do not include the extraction of roots. Polynomials are the simplest example of functional dependence, because they can be evaluated with the aid of the simplest arithmetical operations; there-

fore in the study of more complicated functions attempts are often made to represent them, if only approximately, by polynomials; we shall deal with this aspect in greater detail later.

2. Rational functions. If division is added to the arithmetical operations performed over x and the constants which we discussed above, then we obtain as a result an arbitrary *rational* (generally not an integral) function of x . As an example consider the functions

$$y = \frac{1}{x}, \quad y = \frac{1}{1+x^2}, \quad y = \frac{x^2+1}{x+1}$$

etc. It can be proved by elementary algebra that every rational function can be represented as a relationship of two polynomials, *i.e.* in the form

$$y = \frac{P(x)}{Q(x)}, \quad (1)$$

where $P(x)$ and $Q(x)$ are polynomials. Rational functions, like polynomials, can easily be evaluated for every value of the independent variable x except for those values of x for which $Q(x)=0$ in formula (1); for these latter values of x the rational function, as given by formula (1), remains indefinite; these values correspond to values of x which lie outside the "region of definition" of the function y in accordance with the definition of a region given in § 3; thus if y is given by the formula

$$y = \frac{1}{1-x^2},$$

then its region of definition will be the whole number line with the exception of the points $x=1$ and $x=-1$.

3. General power functions. By this name the following function is known

$$y = x^\alpha,$$

where α is an arbitrary constant. The nature of this function obviously depends on the arithmetical nature of α . If α is an integer then y is a rational function (it is integral when $\alpha \geq 0$). When α is a rational fraction, *e.g.* $\alpha = p/q$ (where p and q are integers and we can always assume that $q > 0$) then

$$x^\alpha = x^{\frac{p}{q}} = \sqrt[q]{x^p}$$

is an *irrational algebraic function* of x (for the operations performed over x include the extraction of a root of an arbitrary degree q).

The values of this function can no longer be evaluated as simply as those of a rational function. This is even more true in the case when α is an irrational number (*i.e.* for example the function $y = x^{\sqrt{2}}$ or $y = x^\pi$); strictly speaking we do not know how to determine such functions; we shall return to this question in §§ 17 and 24.

The region of definition of the function given by the formula $y = x^\alpha$ depends on the nature of the number α . If α is a positive integer then the whole number line serves as its region of definition; but when α is a negative integer or zero, *i.e.* $\alpha \leq 0$, the point $x = 0$ must be excluded from this line. If $\alpha = 1/q$, where q is a positive integer, then the function will be determined for all values of x when q is odd, and only for $x \geq 0$, when q is even. The reader will be able to determine for himself the region of definition of the function x^α when $\alpha = p/q$, where p and q are integers. In cases where α is irrational, its region of definition is the semi-straight line $x > 0$, as we shall learn in § 17.

4. Exponential functions. By this name the following function is known

$$y = a^x,$$

where a is a constant positive number. We shall learn in § 17 that the whole number line serves as the region of definition of this function. We shall learn later some other important properties of this function. The value of y for the given value of x cannot, in this case, be obtained by means of any known finite sequence of operations (with the exception of the trivial case when $a = 1$); the function a^x is not an *algebraic* function but a *transcendental* function*).

*) Strictly speaking the problem is as follows: If the function $y = f(x)$ of the independent variable x is obtained after performing a finite number of algebraic operations, as proved in algebra, then a polynomial $P(x, y)$ of two variables exists so that, identically (*i.e.* for any x) $P[x, f(x)] = 0$. The converse proposition is not true; it may happen that the polynomial P does exist, but the function $f(x)$ cannot be expressed in terms of x by means of a finite number of algebraic operations. It is customary to call the function $f(x)$ an *algebraic* function if a polynomial P exists for this function which possesses the properties mentioned above. Hence the class of algebraic function is wider than the class of functions which can be expressed in terms of a finite number of algebraic operations. Every non-algebraic function is known as a *transcendental* function. The functions a^x , $\log_a x$ (for every $a > 0$, $a \neq 1$), $\sin x$, $\cos x$, $\arcsin x$, $\arccos x$, etc. are examples of transcendental functions.

5. Logarithmic functions. The function

$$y = \log_a x,$$

where a is a constant positive number other than unity, is defined as the inverse of the exponential function. This means that it follows from $y = \log_a x$ that $x = a^y$. To be more exact this means that for every $x > 0$, a single number y exists which satisfies the relationship $a^y = x$; this number y is known as the logarithm of x of the base (or to the base) a and is denoted as $\log_a x$. Like the exponential function the logarithmic function is also a transcendental function; apart from its great theoretical importance it is also very important in calculations; its significance is mainly due to the basic property of this function: $\log_a (\alpha \beta) = \log_a \alpha + \log_a \beta$. The region of definition of a logarithmic function of any base is the semi-straight line $x > 0$.

6. Simpler trigonometrical functions. These functions are the following functions which are well-known from the school course of trigonometry.

$$\begin{array}{lll} y = \sin x, & y = \cos x, & y = \tan x, \\ y = \cot x, & y = \sec x, & y = \operatorname{cosec} x. \end{array}$$

The chief property of these functions is their *periodicity*; $\tan x$ and $\cot x$ have a period π and the remaining four functions a period 2π . The whole number line serves as the region of definition of the functions $\sin x$ and $\cos x$; the functions $\tan x$ and $\sec x$ are defined everywhere except at points of the type

$$y = \left(k + \frac{1}{2}\right) \pi,$$

and the functions $\cot x$ and $\operatorname{cosec} x$ —everywhere except at points of the type

$$y = k\pi,$$

where k in both cases denotes an arbitrary integer.

7. Inverse trigonometrical functions. Generally speaking, the function $\alpha(x)$ is said to be the inverse of the given function $f(x)$ if it follows from $y = \alpha(x)$ that $x = f(y)$. We have seen already that the function $\log_a x$ is the inverse of the function a^x . In this case the inverse function is unique. However it is quite possible for a given function to have several inverse functions; thus the function x^2 evidently has the following inverse functions: $+\sqrt{x}$ and $-\sqrt{x}$, for it follows equally from $y = +\sqrt{x}$ and from $y = -\sqrt{x}$ that $x = y^2$. It is a well-known fact that each one of the simpler trigonometrical functions has an infinite number of inverse functions;

these functions are known as *inverse trigonometrical functions*. Let us consider, for example, the family of functions inverse to the sine. If α is an arbitrary number confined between -1 and $+1$, then an infinite number of values of x exists for which $\sin x = \alpha$; in particular one such value of x can be found between $-\pi/2$ and $+\pi/2$; it is denoted by $\sin^{-1} \alpha$, so that

$$-\frac{\pi}{2} \leq \sin^{-1} \alpha \leq \frac{\pi}{2}, \quad \sin (\sin^{-1} \alpha) = \alpha;$$

it is obvious that the function $\sin^{-1} x$ is the inverse of the function $\sin x$; $\sin^{-1} \alpha$ is one of the angles whose sine is equal to α ; but in such cases, as we know from trigonometry, the general form of an arc, the sine of which is equal to α , is as follows:

$$(-1)^k \sin^{-1} \alpha + k\pi,$$

where k is an arbitrary integer. Hence each one of the functions

$$(-1)^k \sin^{-1} x + k\pi,$$

where k is an arbitrary integer, is a function inverse to the function $\sin x$. The region of definition of all these functions is the line $-1 \leq x \leq 1$. Functions which are the inverse of other simple trigonometrical functions are analysed and defined in a similar way.

The functions considered in sections 1 to 7 include all the *simple* elementary functions. Other elementary functions are obtained from the simple functions either by means of algebraic operations

$$\left[y = \frac{1}{1 + x\pi}, \quad y = 2^x (\cos x - 2 \sin x) \right],$$

or by “superimposition” of functional operations

$$\left[y = \log \cos x, \quad y = \tan \left(1 + 2^{\frac{1}{x^2}} \right) \right],$$

which means that a certain function of the independent variable is first taken, and then another function of this function is taken, etc. As a result of any number of operations of this kind, performed in any order in which the *simple* elementary functions serve as a basis, *all* elementary functions are obtained. We have already said above that we shall study most of the properties of elementary functions later on. Here, as a preliminary review, we have dealt with only a few simple functions.

CHAPTER II

ELEMENTARY THEORY OF LIMITS

§ 7. Infinitesimal Quantities

Variable quantities which we meet in natural phenomena and in technical processes vary in very diverse ways. If we were to begin the study of the various modes of change, one after another, in the order in which we meet them in our practical experiences or in our nature studies, then this would be an unscientific approach to the problem. A botanist does not study all species of plants which happen to catch his eye, but begins by classifying his material, dividing it into groups which resemble each other more or less closely, and only then proceeds with the study of each class of plants as a whole; similarly the mathematician should try to divide all possible types of changes in quantities into more or less extensive classes so as to be able to analyse systematically all the properties which members of a given class have in common. In doing this he always begins with the study of the simpler objects, because in the first place, he learns by experience that the simpler objects of science are, in the majority of cases, of utmost importance in its applications and secondly, it frequently happens in mathematics, that after the simpler cases have been studied, it is possible to break down the more complicated cases to these simple cases and to study them quickly and easily. Thus, when studying equations in algebra we begin with the simplest case, *i.e.* with equations of the first degree with one unknown; this type of equation is most common and more complicated cases can often be broken down to this case.

The history of development of our science has shown that the simplest and the most important type of variable quantities that can subsequently be used in the study of many other quantities, which undergo more complicated changes, are the so-called *infinitesimal quantities*. The leading role of quantities of this type, both in

mathematical theory and in its practical applications is so great that the whole science of changing quantities is even today known by the name of “the analysis of infinitesimal quantities” or “the calculus of infinitesimal quantities”. We therefore begin our study of variables with this type of changes.

Imagine a natural phenomenon, or a technical process in which a certain variable quantity x participates. Generally speaking, in the course of a process, x will increase sometimes and decrease at others. Let us now assume that *the absolute value of x remains infinitesimal during the whole process*. Let us explain in greater detail what this means. Let us assume that we are given a small positive number, for example 0.001. From a certain moment of the process onwards we shall always have $|x| < 0.001$. Assume that we are not satisfied with this degree of smallness and that we want to have $|x| < 0.000001$. In order to achieve this we shall have, generally speaking, to advance the process to a further stage. But from a certain moment onwards we shall always have $|x| < 0.000001$. In general, irrespective of the infinitesimal quantity ε which we might choose, we shall sooner or later reach a moment in our process when $|x| < \varepsilon$ always.

The quantity x , the changes in which (in the given process) display the property described above, is known as an *infinitesimal quantity* (in the given process). Hence we arrive at the following definition:

The quantity x is said to be infinitesimal (in the given process) if an arbitrary constant positive number ε is such that from a certain moment of the process onwards it will always remain $< \varepsilon$.

Example 1. When the temperature is maintained at a constant level, the pressure p of the given mass of gas is inversely proportional to its volume v , i.e.,

$$p = \frac{c}{v}, \quad (1)$$

where c is a positive constant. If we increase the volume of the gas indefinitely, then its pressure decreases; if the process is continued for a long time, i.e. the volume of the gas is sufficiently large, then the pressure of the gas, as can be seen from formula (1), becomes (and when the gas expands further, remains) as small as we please. This means that in the process of the unlimited expansion of the given mass of gas its pressure is an infinitesimal quantity.

Example 2. According to the law of gravitation the sun S attracts the comet K which revolves round it (Fig. 3) with a force

$\frac{k}{r^2}$, where k is a positive constant and r is the distance between the centres of the two heavenly bodies. Let us assume that we are dealing with a comet which only once appears within the reach of the solar system (hyperbolic orbit), after which it retracts indefinitely from it, i.e. the distance r increases indefinitely after the comet has revolved round the sun. It then becomes evident that the force of attraction $\frac{k}{r^2}$ becomes infinitesimal; no matter how small the positive number

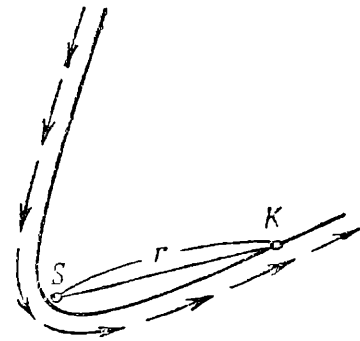


Fig. 3

we choose, this force of attraction will become in the course of this process smaller than ε (and will remain so for ever, i.e. when the comet has retracted from the sun for a sufficiently great distance). This means that the force with which the sun attracts the comet becomes an infinitesimal quantity in the course of the infinite retraction of the comet.

Example 3. In the geometrical progression.

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2^n}, \dots$$

the n 'th term will be as small as we please, provided n is sufficiently great. This means that as n increases indefinitely, $\frac{1}{2^n}$ becomes infinitesimal.

In general, if $0 < \alpha < 1$, then $(1 - \alpha)^n$ is an infinitesimal quantity when n increases indefinitely. In fact from

$$(1 - \alpha)(1 + \alpha) = 1 - \alpha^2 < 1,$$

it follows that

$$1 - \alpha < \frac{1}{1 + \alpha},$$

and therefore, when $n > 0$,

$$(1 - \alpha)^n < \frac{1}{(1 + \alpha)^n};$$

but $(1 + \alpha)^n > 1 + n\alpha$, as can readily be seen by expanding $(1 + \alpha)^n$ by the binomial formula (or by proving it simply by the full method of induction); therefore, the quantity $(1 + \alpha)^n$ becomes as large as we please, provided n is sufficiently large. On the other hand, we can see from our inequality that the quantity $(1 - \alpha)^n$ becomes as small as we please, provided n is sufficiently large, which had to be proved.

Example 4. Fig. 4 represents part of the usual trigonometrical circle of unit radius, so that

$$AD = DC = |\sin x|, \quad \text{arc } AB = \text{arc } BC = |x|.$$

The straight line ADC is shorter than the arc ABC , *i.e.* $2|\sin x| < 2|x|$. Therefore, by decreasing the absolute value of the angle x , we can make the absolute value of the sine as small as we please. This means that in the process of infinitely diminishing the absolute value of an angle, its sine becomes an infinitesimal quantity. This example differs from the preceding example by the fact that $\sin x$ can be both positive and negative; irrespective of this it is an infinitesimal quantity, for according to the definition of an infinitesimal quantity, this type of change is connected only with the *absolute value* of the quantity.

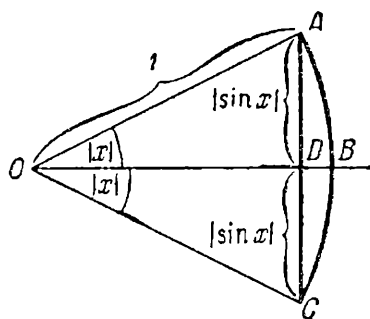


Fig. 4

Example 5. The deflection of the pendulum from the vertical position in Fig. 5 is measured by the angle θ ; it is convenient to regard this angle as positive when deflection occurs to one side (for example, to the right), and as negative when deflection is to the other (to the left). If the pendulum is left to itself

(*i.e.* when its movement is not supported by a spring or weight), then as a result of the friction of the mechanism and the resistance of the air, its amplitude of vibration will continuously decrease. In course of this movement the quantity θ becomes both positive and negative and passes through the zero position each time there is a change of sign. The graph showing the dependence of the angle θ on the time t is schematically represented in Fig. 6 (curve of damped oscillations). In

course of time the height of the waves drops continuously, which indicates a gradual diminution in the amplitude of vibrations. No matter how small the positive number ϵ , sooner or later a moment will be reached when $|\theta| < \epsilon$ always. This means that in the phenomenon under consideration the angle θ is an infinitesimal quantity. We are dealing here with an infinitesimal quantity which changes by acquiring alternately positive and negative values.

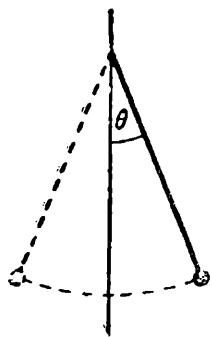


Fig. 5

If the vibration of the pendulum is supported by certain means with the constant expenditure of a form of energy (for example, by using an unwinding spring or descending weight),

then the dependence of the angle θ on time will have the form represented in Fig. 7 (curve of undamped oscillations). In this case the angle θ will no longer be an infinitesimal quantity; it is true to say that in course of time $|\theta|$ becomes an infinitesimal quantity (or even zero); however, no matter how long we wait, we shall never reach the moment after which we shall *always* have $|\theta| < \frac{1}{2}\alpha$, where α is the amplitude of (the undamped) vibrations of the pendulum.

A comparison of the examples shows that infinitesimal quantities can have many diverse modes of change; nevertheless their inclusion in a single class presents, as we shall see on many occasions later on, a very convenient method of investigation.

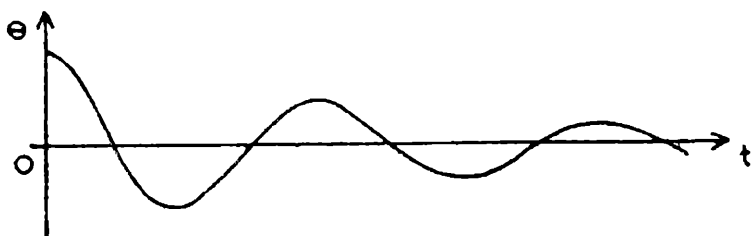


Fig. 6

Note. The term “infinitesimal quantity” has been so well-established historically that it is very difficult to replace it by any other term without causing chaos in scientific terminology. However, this term is rather unfortunate and from the pedagogical point of view it conceals a danger about which the student must be warned. The word “infinitesimal” sounds as if it were intended to indicate the dimension of the quantity in question and the student frequently connects the term “infinitesimal” with the concept of a “very small” quantity or a “negligible quantity”. This is quite incorrect. The term “infinitesimal” describes, by its definition, *not the dimensions* of a quantity, but the *character of its change*. It would, of course, be more correct to call such quantities not “infinitesimal” but “indefinitely decreasing.”

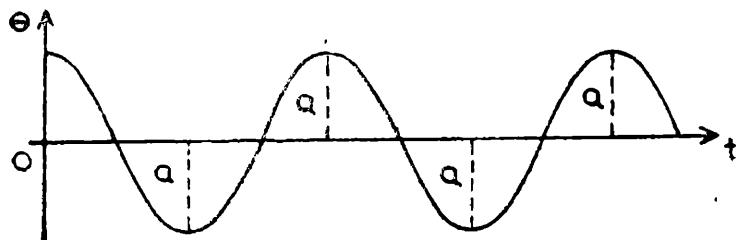


Fig. 7

§ 8. Operations with Infinitesimal Quantities

The wide application of infinitesimal quantities to changes occurring in the world is made considerably easier by the fact that as a result of the simpler algebraic operations with infinitesimal quantities, other infinitesimal quantities are obtained. We shall now formulate this property in several simple theorems.

Theorem 1. *The algebraic sum of a constant number of infinitesimal quantities is an infinitesimal quantity.*

Proof. Let $s = x_1 \pm x_2 \pm \dots \pm x_n$, where x_1, x_2, \dots, x_n are infinitesimal quantities and the number n is a constant. It is to be proved that the quantity s is infinitesimal.

Let ε be an arbitrary positive number; in that case the number ε/n will also be positive and constant. Since x_1 is an infinitesimal quantity, we shall reach a moment in our process after which we shall always have

$$|x_1| < \frac{\varepsilon}{n};$$

the same applies to x_2 which is also an infinitesimal quantity and will also, in the course of the process, reach a point after which we shall always have

$$|x_2| < \frac{\varepsilon}{n};$$

the same also applies to x_3, x_4, \dots, x_n . Hence the absolute value of every term of the sum s sooner or later reaches a moment after which it always remains smaller than ε/n ; the moments when this inequality is reached will, generally speaking, vary for different terms. However the number of these moments is equal to the number of terms n and among them the *latest* moment can be found; from this latest moment onwards *all* the n inequalities will always be satisfied:

$$|x_1| < \frac{\varepsilon}{n}, |x_2| < \frac{\varepsilon}{n}, \dots, |x_n| < \frac{\varepsilon}{n},$$

and therefore, the inequality obtained as a result of term-to-term addition will also be satisfied

$$|x_1| + |x_2| + \dots + |x_n| < n \cdot \frac{\varepsilon}{n} = \varepsilon,$$

and therefore,*)

$$|s| = \left| \sum_{k=1}^n \pm x_k \right| \leq \sum_{k=1}^n x_k < \varepsilon.$$

We have thus shown that no matter how small the positive number ε is, a point will be reached in the process after which we shall always have $|s| < \varepsilon$. This means that s is an infinitesimal quantity. Theorem 1 is thus proved.

In order to prove the other theorems we must introduce one more concept, which in future, will prove to be of great importance. We say that a certain quantity y which participates in a given process is *limited* (in this process), if there exists a positive number C and there is a moment in this process after which we always have $|y| < C$. This definition somewhat resembles the definition of an infinitesimal quantity; however, there is an essential difference: an infinitesimal quantity should in the course of the process become and remain (by its absolute value) smaller than an *arbitrary* positive number and a limited quantity should be smaller than *at least one* positive number. It therefore, follows that *every infinitesimal quantity is also a limited small quantity*. But the converse statement is not true. Thus the distance of the earth (or any other planet) from the sun is, evidently, a limited quantity, but it is not an infinitesimal quantity. Another example: if the number x increases continuously, then

*) Here and elsewhere we are using the abbreviated notation of a sum which is generally accepted in mathematics:

$$a_m + a_{m+1} + \dots + a_n = \sum_{k=m}^n a_k,$$

where m and n ($m < n$) are arbitrary integers, a_k ($m \leq k \leq n$) are arbitrary numbers; thus for example

$$\sum_{k=3}^8 \frac{1}{k^2}$$

denotes the sum

$$\frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} + \frac{1}{8^2}.$$

The inequality in the text is based on the well-known algebraic rule: *the absolute value of an algebraic sum does not exceed the sum of the absolute values of the terms.*

$\sin x$ is a limited quantity (because we always have $|\sin x| < 2$), but it is not an infinitesimal quantity (because, in any case, we obtain the value $|\sin x| = 1$ again and again). Therefore the concept of a limited quantity is wider (more general) than the concept of an infinitesimal quantity.

Theorem 2. *The product of an infinitesimal quantity and a limited quantity is an infinitesimal quantity.*

Proof. Let us assume that in a certain process x is an infinitesimal quantity and y is a limited quantity and let ε be an arbitrary positive constant. It follows from the definition of the limited quantity y that a number c exists which is such that from a certain moment of our process onwards we always have $|y| < c$. On the other hand, from another moment onwards (owing to the infinitesimality of x), we have $|x| < \varepsilon/c$. Hence after the occurrence of the latter of the two moments, both inequalities $|y| < c$ and $|x| < \varepsilon/c$ will be satisfied; hence the inequality obtained as a result of the term-by-term multiplication

$$|xy| = |x| \cdot |y| < \frac{\varepsilon}{c} \cdot c = \varepsilon.$$

will also be satisfied.

The fact that the number ε can be chosen arbitrarily means that xy is an infinitesimal quantity, which proves the theorem 2.

Corollary 1. *The product of an infinitesimal quantity and a constant is an infinitesimal quantity.*

We saw above that every infinitesimal quantity is also limited; we therefore have

Corollary 2. *The product of two infinitesimal quantities is an infinitesimal quantity.*

By means of the method of induction this proposition can be extended to include any number of factors. If x_1, x_2, x_3 are infinitesimals, then as a result of the corollary 2, the product $x_1 x_2$ is also infinitesimal; this, as a result of the same corollary 2, shows that $(x_1 x_2) x_3 = x_1 x_2 x_3$ must also be an infinitesimal quantity. From three factors we can go on to four factors, etc. We therefore have

Corollary 3. *The product of any constant number of infinitesimal quantities is an infinitesimal quantity.*

And, in particular,

Corollary 4. *An arbitrary positive integral power of an infinitesimal quantity is an infinitesimal quantity.*

We can thus see that the operations of addition, subtraction and raising to an integral positive power performed any number of times and in any order over infinitesimal quantities result in other infinitesimal quantities. It is not by chance that these operations do not include *division*. The *quotient* of two infinitesimal quantities may not be an infinitesimal quantity. In fact, let us assume that the quantity x which is involved in a certain process is infinitesimal. It follows from corollary 4 that the quantity x^2 will also be an infinitesimal quantity in this process. Let us assume, for the sake of simplicity, that x is never zero; in this event each one of the three fractions

$$\frac{x^2}{x}, \frac{x}{x}, \frac{x}{x^2}$$

represents the quotient of two infinitesimal quantities. The first fraction is equal to x and is therefore infinitesimal; the second fraction is equal to unity and is therefore limited, but not infinitesimal; finally the third fraction is equal to $1/x$; thus in our process, as $|x|$ tends to become infinitesimal, $|1/x| = 1/|x|$ tends to become as large as we please and therefore $1/x$, i.e., our third fraction, is not only not infinitesimal, but it is not even limited.

If the quantity x which participates in a process is equal to zero throughout this process, then $|x|$ is at every moment of this process smaller than an arbitrary positive number ϵ . It follows from the definition of an infinitesimal quantity that x must be infinitesimal:

A quantity which is equal to zero through a process is an infinitesimal quantity in this process.

§ 9. Infinitely Large Quantities

We shall now study another aspect of changes in quantities which is the opposite of infinitesimality.

A quantity x is said to be an infinitely large quantity in a given process if from a certain moment of this process onwards it becomes greater than a positive number A , which can be as large as we please, so that we have always $|x| > A$.

Infinite greatness, like infinitesimalness, is thus fully defined by the behaviour of the absolute value of the given quantity, quite apart from its sign, so that together with x , the quantity $|x|$ must also be infinitely large. With regard to an infinitely large quantity the same warning must be given as that given in § 7: infinite greatness does not tell us the dimensions of the quantity studied and tells us only of the manner in which it changes; it is therefore incorrect to connect the concept of an "infinitely large quantity" with the concept of a quantity with very large dimensions.

Example 1. The distance r from the sun to the comet in the example 2, § 7, is an infinitely large quantity in the process of movement of the comet.

Example 2. If the acute angle x approaches a right angle, then $\tan x$ in this process is an infinitely large quantity. The same thing will take place when the *obtuse* angle x approaches a right angle (in this case $\tan x$ is negative).

Example 3. If the number n increases indefinitely, then $(-1)^n 2^n$ is an infinitely large quantity (since $|(-1)^n 2^n| = 2^n$). It can be seen from this example that an infinitely large quantity, like an infinitesimal quantity, can change its sign in the course of the process an infinite number of times.

Example 4. In the example 3, § 7, we saw that for every constant $\alpha > 0$, the quantity $(1 + \alpha)^n$ is an infinitely large quantity when $n \rightarrow \infty$.

Let us now consider the operation with infinitely large quantities. The sum of two infinitely large quantities need not necessarily be an infinitely large quantity, as can be seen from the following simple example: if x is an infinitely large quantity, then as we saw, $-x$ will also be an infinitely large quantity; the sum of these two quantities is always equal to zero, *i.e.* it is an infinitesimal quantity.

We have, however, the following important theorems:

Theorem 1. *The sum of two quantities, one of which is infinitely large and the other is a limited quantity, is an infinitely large quantity.*

Proof. Let us assume that in the given process x is infinitely large and y is a limited quantity. A positive number C exists which is such that from a certain moment of our process onwards we always have $|y| < C$; let A be an arbitrary positive number; owing to the fact that x is infinitely large, there will be another moment in our process

after which we always have $|x| > A + C$. Hence by choosing the latter of the two moments, we shall always have after that instant:

$$|x| > A + C, \quad |y| < C,$$

hence *)

$$|x + y| \geq |x| - |y| > A + C - C = A.$$

Owing to the fact that the number A is arbitrarily large, this proves that $x + y$ is an infinitely large quantity.

If, as we saw above, the addition of infinitely large quantities does not always lead to infinitely large quantities, then on the other hand, the multiplication of infinitely large quantities follows the same rule as the multiplication of infinitely small quantities.

Theorem 2. *The product of two infinitely large quantities is an infinitely large quantity.*

Proof. The reader is by now well-acquainted with the arguments used for proving theorems of this kind; we can, therefore, treat the proof more briefly. If x_1 and x_2 are infinitely great in the given process and if A is an arbitrary positive number, then from a certain moment of the process onwards $|x_1| > \sqrt{A}$, and from another moment onwards $|x_2| > \sqrt{A}$; but from the latter of the two moments onwards $|x_1 x_2| = |x_1| \cdot |x_2| > A$, which proves the theorem.

From this, in the same way as with infinitesimal quantities, we obtain with the aid of induction :

Corollary. *The product of an arbitrary constant number of infinitely large quantities is an infinitely large quantity.*

The following proposition connects the concept of an infinitely large quantity with the concept of an infinitely small quantity :

Theorem 3. *If x is an infinitesimal quantity which is never zero, then $1/x$ is an infinitely large quantity; conversely, if x is an infinitely large quantity which is never zero, then $1/x$ is an infinitesimal quantity.*

To prove this it is sufficient to note that the inequality $|x| < \varepsilon$ is equivalent to the inequality $1/|x| > 1/\varepsilon$, and if the number ε is as small as we please, then the number $1/\varepsilon$ is as large as we please.

*) We are using here a rule which is well-known in elementary algebra : the absolute value of a sum is not less than the difference of the absolute values of its terms.

§ 10. Quantities which tend to Limits

In the above sections we have dealt with some of the simple types of changes in quantities, *i.e.* we have considered quantities which decreased indefinitely and other quantities which increased indefinitely and which are known as infinitesimal and infinitely large quantities respectively. Following our scheme, we shall now consider the next large class of a type of change and in doings, so we shall find the concept of infinitesimal quantities very useful.

In practice and in natural phenomena it happens frequently that the variable quantity x tends to come infinitely close to a certain constant a , so that in the course of the process the absolute value of the difference between these quantities becomes infinitesimal; in such cases it is said that the quantity x has a *limit* a in the given process or that it *tends to* a . This is denoted as follows: $\lim x = a$, or $x \rightarrow a$. the two forms of notation are equivalent. The word *lim* is made up of the first three letters of the latin word *limes* which means limits or boundary; but the word should be read in English, *i.e.* "limit".

It is obvious that the quantity x cannot, in this case, have two different limits: in fact, if $x \rightarrow a_1$ and $x \rightarrow a_2$, then the absolute values of the quantities $x - a_1$ and $x - a_2$ become and remain infinitesimal in the given process; hence their difference, *i.e.* the absolute values of the constant $a_2 - a_1$ must, in the course of the process, also become and remain infinitesimal, which is only possible when $a_2 = a_1$.

As we have just said above, the relationship $\lim x \rightarrow a$ (or $x \rightarrow a$), where a must be a constant, means that the absolute value of the difference $x - a$ becomes and remains in the course of the given process as small as we please, *i.e.* it is smaller than an arbitrary constant positive number. But, by definition, a quantity which changes in this manner is an infinitesimal quantity. We can therefore say that

The quantity x tends in the given process to a constant quantity a (or, which is equivalent, it has as its limit the constant quantity a), provided the difference $x - a$ is an infinitesimal quantity in this process.

Example 1. A heated body (temperature T_1) is immersed in water (temperature $T_2 < T_1$). The body cools down gradually (T_1 falls) and the surrounding water warms up (T_2 rises); both quantities T_1 and T_2 thus tend indefinitely towards an average temperature T ($T_2 < T < T_1$), so that, in the course of the process the differences $T_1 - T$ and $T_2 - T$ become infinitesimal.

We thus have

$$\lim T_1 = T, \quad \lim T_2 = T,$$

or

$$T_1 \rightarrow T, \quad T_2 \rightarrow T.$$

Example 2. A coin is thrown n times in succession and after each throw it is noted whether the head or tail turns upwards. Let us assume that after the coin is thrown n times, the head appears m times at the top; as n increases m , increases as well. Experience shows that when the coin is geometrically regular and physically homogeneous, then, provided it is thrown a great many times, the head will appear at the top half the number of times, *i.e.* the relationship m/n tends to $1/2$; we can take it to be proved empirically that the absolute value of the difference

$$\frac{m}{n} - \frac{1}{2},$$

when n increases indefinitely (by becoming both positive and negative), remains in the end as small as we please, *i.e.* this difference becomes an infinitesimal quantity as the number of times the coin is thrown increases indefinitely. Therefore in our process

$$\lim \frac{m}{n} = \frac{1}{2} \quad \text{or} \quad \frac{m}{n} \rightarrow \frac{1}{2}.$$

Example 3. If the quantity x in a certain process is infinitesimal, then the quantity $y = a + bx + cx^2$, where a , b and c are constants, tends to the limit a in this process. In fact, $y - a = bx + cx^2$ and if x is infinitesimal, then the theorem in § 8 enables us to say that the quantity $bx + cx^2$ is also infinitesimal.

Example 4. If in a certain process the quantity x is infinitesimal, then $\cos x$ tends to unity as its limit. In fact, from a certain moment of the process onwards $|x| < \pi/2$, *i.e.* the angle x is an acute angle and $\cos x > 0$. It follows from the relationship $1 - \cos^2 x = \sin^2 x$ that

$$0 \leq 1 - \cos x = \frac{\sin^2 x}{1 + \cos x} < \sin^2 x,$$

and owing to the fact that apart from x , $\sin x$ is also an infinitesimal quantity (example 4, § 7), it follows that $\sin^2 x$ (corollary 4 of theorem 2, § 8) and the quantity $1 - \cos x$ are also infinitesimal and are

confined between zero and the infinitesimal value $\sin^2 x$; but in our process this means that

$$\cos x \rightarrow 1.$$

Example 5. Let us prove that for every constant $a > 0$ the quantity $\sqrt[n]{a}$ tends to unity as its limit, provided n increases indefinitely. In fact let it be given arbitrarily that $\varepsilon > 0$; we know (example 3, § 7) that as n increases indefinitely, the quantity $(1 - \varepsilon)^n$ is an infinitesimal quantity and the quantity $(1 + \varepsilon)^n$ is an infinitely large quantity; therefore provided n is sufficiently large, we have

$$(1 - \varepsilon)^n < a < (1 + \varepsilon)^n,$$

hence $1 - \varepsilon < \sqrt[n]{a} < 1 + \varepsilon,$

or $|\sqrt[n]{a} - 1| < \varepsilon,$

which proves our proposition.

The above examples show us that the way in which a variable quantity tends to its limit can be very diverse in character. Thus in example 1 the temperature T_1 tends to its limit T by decreasing continuously; on the other hand, the temperature T_2 (in the same example) tends to this same limit T by increasing continuously. In example 2 (the experiment with throwing a coin) theory and practice show us that by increasing the number of times the coin is thrown, the "fraction of heads" m/n becomes greater and smaller, (and sometimes equal to) $1/2$; we are dealing here with a quantity which increases and decreases in the process under consideration while it tends towards its limit.

In spite of the fact that quantities show great differences in behaviour when tending towards their limit, they also share many properties in common; this makes it possible to include them in the same class. We shall now study some of their properties.

Theorem 1. *A quantity which tends to a limit in a given process is a limited quantity in this process.*

Proof. Let us assume that in a certain process $x \rightarrow a$. In this case the difference $x - a$ is infinitesimal and, therefore, from a

certain moment of the process onwards $|x - a| < 1$; hence owing to the fact that $x = a + (x - a)$, we have

$$|x| \leq |a| + |x - a| < |a| + 1$$

This inequality, on the right-hand side of which stands a certain constant positive number, is satisfied from a certain moment of the process onwards; but this means that the quantity x is limited in this process.

Theorem 2. *If in a certain process $x \rightarrow a$ and $a > 0$, then from a certain moment of the process onwards we shall always have $x > 0$.*

In other words, if a given quantity has a positive limit, then from a certain moment of the process onwards the quantity itself will be positive.

Proof. Let b be an arbitrary positive number smaller than a ($0 < b < a$). Owing to the fact that the difference $x - a$ is infinitesimal, we shall have from a certain moment of the process onwards

$$|x - a| < b;$$

since $x = a + (x - a)$, we have from that moment onwards

$$x \geq a - |x - a| > a - b > 0,$$

which had to be proved.

Corollary 1. *If in a certain process $x \rightarrow a$ and $a < 0$, then from a certain moment of the process onwards we shall always have $x < 0$.*

Corollary 2. *If $x \rightarrow a$, and from a certain moment of the process onwards $x \geq 0$, then $a \geq 0$. If from a certain moment of the process onwards $x \leq 0$, then $a \leq 0$.*

The proof of both these corollaries is so obvious that we shall not give it here.

Let us now assume that in a certain process $x \rightarrow 0$. This, as we know, is equivalent to the fact that $x - 0 = x$ is an infinitesimal quantity; we thus arrive at the following proposition:

Theorem 3. *Every infinitesimal quantity has zero as its limit and, conversely, every quantity which tends to zero is infinitesimal.*

This theorem is very important. It shows that infinitesimal quantities which we considered earlier are a particular case of quantities which tend to a limit. On the other hand, infinitely large quantities cannot tend to a limit; this follows from theorem 1, for, evidently, an infinitely large quantity cannot be limited.

We have finally :

Theorem 4. *Every constant a is its own limit.*

In order to prove this it is sufficient to say that the relationship $a \rightarrow a$ is equivalent to the fact that $a - a$ must be an infinitesimal quantity; but $a - a = 0$ and the constant zero is, as we know, always an infinitesimal quantity (cf. end of § 8).

Other properties of quantities which tend to limits are connected with operations with these quantities; we shall deal with these in the following paragraph.

§ 11. Operations with quantities which tend to limits

Theorem 1. *If in a certain process $x_1 \rightarrow a_1$, $x_2 \rightarrow a_2$, ..., $x_n \rightarrow a_n$, then*

$$x_1 \pm x_2 \pm \dots \pm x_n \rightarrow a_1 \pm a_2 \pm \dots \pm a_n.$$

This theorem is frequently formulated as follows: the limit of an algebraic sum (with a constant number of terms) is equal to the algebraic sum of limits; this formulation is even more obvious if the theorem is written down in its equivalent form:

$$\lim (x_1 \pm x_2 \pm \dots \pm x_n) = \lim x_1 \pm \lim x_2 \pm \dots \pm \lim x_n.$$

It is only necessary to remember that it is assumed that each term has a limit; this is the necessary requirement of this theorem; on the other hand the existence of a limit of the algebraic sum as a whole is then no longer assumed but maintained (and, of course, proved). The full (but rather lengthy) formulation of theorem 1 should read as follows: *If in a certain process every quantity x_i ($1 \leq i \leq n$) has a limit, then the algebraic sum of these quantities also has a limit and this limit of the algebraic sum is equal to the algebraic sum of the limits of terms* This note also refers to all subsequent theorems of this kind.

Proof. It follows from the assumptions made with regard to this theorem that in the given process all the differences

$$x_1 - a_1 = \alpha_1, \quad x_2 - a_2 = \alpha_2, \quad \dots, \quad x_n - a_n = \alpha_n$$

are infinitesimal quantities. It follows from theorem 1, § 8 that their algebraic sum $\alpha_1 \pm \alpha_2 \pm \dots \pm \alpha_n$ is also an infinitesimal quantity but this algebraic sum is, evidently, equal to

$$(x_1 \pm x_2 \pm \dots \pm x_n) - (a_1 \pm a_2 \pm \dots \pm a_n);$$

from which it follows directly that

$$x_1 \pm x_2 \pm \dots \pm x_n \rightarrow a_1 \pm a_2 \pm \dots \pm a_n,$$

and the present theorem is thus proved.

Theorem 2. *If in a certain process $x_1 \rightarrow a_1, x_2 \rightarrow a_2, \dots, x_n \rightarrow a_n$, then*

$$x_1 x_2 \dots x_n \rightarrow a_1 a_2 \dots a_n.$$

Proof. To begin with, let us prove this theorem for two factors ($n = 2$). Let us assume that $x_1 - a_1 = \alpha_1$, $x_2 - a_2 = \alpha_2$ and that α_1 and α_2 are infinitesimal quantities. Hence

$$x_1 = a_1 + \alpha_1, \quad x_2 = a_2 + \alpha_2,$$

$$x_1 x_2 = a_1 a_2 + a_1 \alpha_2 + a_2 \alpha_1 + \alpha_1 \alpha_2,$$

$$x_1 x_2 - a_1 a_2 = a_1 \alpha_2 + a_2 \alpha_1 + \alpha_1 \alpha_2.$$

On the right-hand side of the last equation all the three terms are infinitesimal (the first two as a result of corollary 1, and the last as a result of corollary 2 of theorem 2, § 8); it follows from theorem 1, § 8 that the right-hand sides and therefore also the left-hand sides are infinitesimals. But the infinitesimality of the difference $x_1 x_2 - a_1 a_2$ means that $x_1 x_2 \rightarrow a_1 a_2$; hence we prove the theorem in the case when $n = 2$. It is not difficult to prove the theorem for $n = 3$, and later for $n = 4$, etc.; thus for example if $x_3 \rightarrow a_3$ and the theorem is already proved for $n = 2$. Hence

$$\begin{aligned} \lim (x_1 x_2 x_3) &= \lim [(x_1 x_2) x_3] = \lim (x_1 x_2) \lim x_3 = \\ &= \lim x_1 \lim x_2 \lim x_3, \end{aligned}$$

which proves the theorem 2 for $n = 3$.

Theorem 3. *If in a certain process $x \rightarrow a$ and k is a constant, then $kx \rightarrow ka$.*

Owing to the fact that $k \rightarrow k$ as a result of the theorem 4, § 10, it follows that theorem 3 is an immediate corollary of theorem 2. Theorem 3 can also be formulated as follows:

$$\lim (kx) = k \lim x,$$

as a result of which this theorem can also be formulated as follows: *a constant factor can be taken outside the limit sign.*

Theorem 4. *If in a certain process $x \rightarrow a$ and n is an arbitrary constant natural number, then $x^n \rightarrow a^n$.*

This theorem is obviously a particular case of theorem 2.

Theorem 5 follows from the theorems 4, 3 and 1.

Theorem 5. *If $P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n$ is an arbitrary polynomial of x and if in a certain process $x \rightarrow a$, then in this process $P(x) \rightarrow P(a)$.*

Example. Let $P(x) = 2x^3 - 4x^2 + 5x - 12$. If in a certain process $x \rightarrow 2$, then

$$P(x) \rightarrow P(2) = -2.$$

So far we have only studied the operations of addition, subtraction, multiplication and raising to a power with a constant natural index as applied to quantities which tend to limits. We shall now go on with theorems connected with division.

Theorem 6. *If in a certain process $x \rightarrow a$ and $a \neq 0$, then $1/x \rightarrow 1/a$ in this process.*

Proof. To begin with, it follows from theorem 2, § 10 (or from its first corollary) that because $a \neq 0$, we have from a certain moment onwards $x \neq 0$, so that $1/x$ is not devoid of meaning. Furthermore, owing to the fact that $x - a$ is an infinitesimal quantity, we shall have from a certain moment onwards $|x - a| < \frac{1}{2}|a|$ and therefore

$$|x| = |a + (x - a)| \geq |a| - |x - a| > |a| - \frac{1}{2}|a| = \frac{1}{2}|a|,$$

hence

$$\frac{1}{|x|} < \frac{2}{|a|},$$

which means that

$$\frac{1}{|ax|} < \frac{2}{a^2}.$$

This inequality shows that $1/ax$ is a bounded quantity in our process. Therefore the quantity

$$\frac{1}{x} - \frac{1}{a} = \frac{1}{ax}(a - x)$$

as a result of theorem 2, § 8 is an infinitesimal quantity, which means that $1/x \rightarrow 1/a$. Theorem 6 is thus proved.

Theorem 7. *If in a certain process $x_1 \rightarrow a_1$, $x_2 \rightarrow a_2$ and if $a_2 \neq 0$, then $x_1/x_2 \rightarrow a_1/a_2$.*

The proof of this theorem follows directly from the theorems 6 and 2 ($n = 2$).

Corollary. If $P(x)$ and $Q(x)$ are two arbitrary polynomials of x and in a certain process $x \rightarrow a$ and $Q(a) \neq 0$, then

$$\frac{P(x)}{Q(x)} \rightarrow \frac{P(a)}{Q(a)}.$$

It follows from the fact that $x \rightarrow a$ and from theorem 5 that $P(x) \rightarrow P(a)$, $Q(x) \rightarrow Q(a)$; therefore the relation to be proved is a particular case of theorem 7.

Theorem 7 enables us to express the limit of a fraction in terms of the limits of the numerator and the denominator of this fraction in all cases when these limits exist and when the limit of the denominator is other than 0. If $a_2 = 0$, then this theorem cannot be used for the study of the fraction x_1/x_2 . It can be readily seen that when $a_2 = 0$, the fraction x_1/x_2 can only have a limit provided $a_1 = 0$; since $x_1 = (x_1/x_2)x_2$, then provided $\lim x_2 = 0$ and there is a $\lim x_1/x_2 = b$, we have from theorem 2 ($n = 2$):

$$a_1 = \lim x_1 = \lim \frac{x_1}{x_2} \lim x_2 = b \cdot 0 = 0.$$

Hence we arrive at the following conclusion.

Theorem 8. If the denominator of a fraction is infinitesimal, then the fraction can only have a limit provided its numerator is also infinitesimal.

In this case the given fraction represents a quotient of two infinitesimal quantities. Such a quotient, as we saw in § 8, can change in very diverse ways; therefore every case must be studied on its own merit. We must, however, emphasize that the analysis of the mode of change which is characteristic of one or other ratio of two infinitesimal quantities is, as we shall see later, one of the most important problems in mathematical analysis. At the end of this section we shall consider a definite case of a problem of this kind.

Before doing this, however, we shall establish two more propositions which are very important in the assessment and the practical evaluation of limits of variable quantities.

Theorem 9. If $x \rightarrow a$, $y \rightarrow b$ and beginning from a certain moment of the process $x \geq y$ implies $a \geq b$.

For proving this theorem, it is only necessary to apply corollary 2 of theorem 2, § 10 to the difference $x - y$.

Theorem 10. *If starting from some moment of the process we always have*

$$x_1 \leq x \leq x_2, \quad (1)$$

and if the quantities x_1 and x_2 tend in this process to the same limit a , then a is the limit of x .

Proof. It follows from (1) that

$$0 \leq x - x_1 \leq x_2 - x_1,$$

hence

$$|x - x_1| \leq |x_2 - x_1|; \quad (2)$$

but $\lim (x_2 - x_1) = \lim x_2 - \lim x_1 = a - a = 0$, so that $x_2 - x_1$ is infinitesimal; it follows from (2) that $x - x_1$ is also infinitesimal, so that $x - x_1 \rightarrow 0$; but

$$x = x_1 + (x - x_1);$$

since $x_1 \rightarrow a$ and $x - x_1 \rightarrow 0$, therefore $x \rightarrow a + 0 = a$, which was to be proved.

The importance of theorem 10 is due to the fact that in definite cases the quantity x , the limit of which we are trying to find, has sometimes a complicated form difficult to analyse, which makes our problem very difficult; it is then possible to say that x always lies between two other quantities x_1 and x_2 which have a considerably simpler form: if we also succeed in showing that the quantities x_1 and x_2 tend to the same limit a , then it follows from theorem 10 that $x \rightarrow a$; we are thus able to find the limit of x without having to analyse directly its complicated expression. Let us now consider

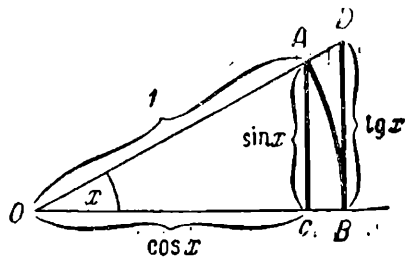


Fig. 8.

a definite example of this kind which we shall be able to use as one of the most important examples in calculating the limit of the ratio of two infinitesimal quantities.

Problem. *Let x be an infinitesimal quantity in a certain process; it is required to prove that $\sin x / x$ has a limit and to find this limit*.*

* We are assuming for the sake of simplicity that x does not vanish in this process.

Solution. Fig. 8 represents part of the usual trigonometrical circle of unit radius. The area of the triangle OAC is evidently equal to $\frac{1}{2} \sin x \cos x$ and is smaller than the area of the circular sector OAB which is equal to $\frac{1}{2} x$; the latter is smaller than the area of the triangle ODB which is equal to

$$\frac{1}{2} \tan x = \frac{1}{2} \frac{\sin x}{\cos x}.$$

Hence

$$\frac{1}{2} \sin x \cos x < \frac{1}{2} x < \frac{1}{2} \frac{\sin x}{\cos x},$$

and therefore

$$\cos x < \frac{x}{\sin x} < \frac{1}{\cos x}.$$

But when $x \rightarrow 0$, we have (example 4, § 10) $\cos x \rightarrow 1$, which means that according to theorem 6, $1/\cos x \rightarrow 1$. Both the right and left-hand sides of the above inequalities tend to unity when $x \rightarrow 0$. Therefore, according to theorem 10, we can conclude that

$$\frac{x}{\sin x} \rightarrow 1,$$

and therefore (by applying theorem 6 again)

$$\lim \frac{\sin x}{x} = 1,$$

when $x \rightarrow 0$. For the sake of simplicity we are assuming that $x > 0$; but owing to the fact that the value of $\sin x/x$ remains unaltered when x is replaced by $-x$, the result remains valid for every approximation of x to zero.

The above result is of the utmost importance in finding limits of quantities, the expressions of which include trigonometrical functions. Thus if x is an infinitesimal quantity, then the numerator and the denominator of the fraction

$$y = \frac{1 - \cos x}{x^2}$$

are also infinitesimal; owing to the fact that

$$1 - \cos x = \frac{\sin^2 x}{1 + \cos x};$$

we have

$$y = \frac{\sin^2 x}{1 + \cos x} \cdot \frac{1}{x^2} = \frac{1}{1 + \cos x} \left(\frac{\sin x}{x} \right)^2;$$

but, when $x \rightarrow 0$, we have :

$$\cos x \rightarrow 1, \quad \frac{\sin x}{x} \rightarrow 1,$$

which means that

$$y = \frac{1 - \cos x}{x^2} \rightarrow \frac{1}{2}.$$

§ 12. Infinitesimal and infinitely large quantities of different orders

We shall return for a short time to infinitesimal and infinitely large quantities in order to somewhat supplement their theories.

Let us assume that in a certain process two infinitesimal quantities x and y are involved and that we wish to compare them with respect to the rate at which they diminish. To do this let us consider the ratio y/x (we shall assume for the sake of simplicity that at least, from a certain moment of our process onwards, x does not vanish so that the quantity y/x is at no time devoid of meaning). It may happen — and we saw examples of this kind in § 8 — that the ratio y/x is itself infinitesimal; this evidently means that in our process y is an infinitesimal quantity and that it is not only small in itself but it is also small *in comparison with the infinitesimal x* , so that provided the process is sufficiently advanced, $|y|$ will only comprise a negligible part of $|x|$; this will be so, for example, when $y = x^2$. In this case we say that y is an *infinitesimal quantity of a higher order* as compared to x . Conversely, x , in comparison with y , has a *lower degree of smallness*.

Let us now assume that the ratio y/x is an infinitely large quantity in the given process; it follows from theorem 3, § 9 that the

reciprocal ratio x/y will, in this case, be infinitesimal and therefore x has a higher degree of smallness as compared to y (and y has a lower degree of smallness as compared to x).

Let us finally consider the case when in a given process the ratio y/x of two infinitesimal quantities can neither decrease indefinitely nor increase indefinitely, but its absolute value remains confined between two positive boundaries; this means that two positive constant numbers a and b exist so that from a certain moment of our process onwards we have

$$a < \left| \frac{y}{x} \right| < b.$$

It is obvious that this means that in the given process neither of the quantities $|x|$ and $|y|$ can, by decreasing, considerably outdo one another. In this case it is said that *the quantities x and y have the same order of smallness*, or they are *infinitesimal quantities of the same order*. In particular this occurs when in the given process the ratio y/x tends to a limit a other than zero; in fact, if this is so, then no matter how small $\varepsilon > 0$, we shall have from a certain moment onwards

$$|a| - \varepsilon < \left| \frac{y}{x} \right| < |a| + \varepsilon,$$

where, provided ε is sufficiently small, the number $|a| - \varepsilon$ and $|a| + \varepsilon$ are positive constants.

It is evident that in the course of their variation the infinitesimal quantities x and y are particularly close to one another when

$$a = \lim \left(\frac{y}{x} \right) = 1;$$

and in this case x and y are said to be *equivalent* infinitesimal quantities. The equivalence of infinitesimal quantities x and y is denoted as $x \sim y$. We have shown at the end of the previous paragraph that the infinitesimal quantities x and $\sin x$ are mutually equivalent. The concept of equivalence of infinitesimal quantities is very important in the evaluation of limits; its importance is based on the following proposition :

Theorem 1. *If x and y are equivalent infinitesimal quantities and z is a third quantity which is involved in this process, then it follows from $xz \rightarrow a$ that $yz \rightarrow a$.*

In other words, when a quantity tends towards a limit and if one infinitesimal factor is replaced by another equivalent infinitesimal quantity, then the quantity which has been changed in this way will tend to the same limit.

To prove theorem 1 it is sufficient to say that it follows from

$$yz = \frac{y}{x} \cdot xz$$

that

$$\lim (yz) = \lim \left(\frac{y}{x} \right) \lim (xz) = 1 \cdot a = a.$$

With the aid of theorem 1 it is frequently possible to replace individual infinitesimal factors of an expression by other equivalent but simpler, infinitesimal quantities; this is very useful in evaluating limits; it is thus possible to simplify the solution of a given problem. Thus in the solution of the last problem in the previous paragraph we could have replaced in the expression

$$y = \frac{\sin^2 x}{1 + \cos x} \cdot \frac{1}{x^2}$$

$\sin^2 x$ by x^2 , since $\sin x \sim x$, *i.e.* we could have written simply

$$\lim y = \lim \frac{1}{1 + \cos x} = \frac{1}{2}.$$

The same equivalence of $\sin x \sim x$ makes it possible, for example, to find from infinitesimal quantity x ,

$$\lim \frac{\sin x}{x^3 + 3x} = \lim \frac{x}{x(x^2 + 3)} = \lim \frac{1}{x^2 + 3} = \frac{1}{3}.$$

In the above we considered cases of two infinitesimal quantities x and y which take part in a given process such that: (1) one has a higher degree of smallness as compared with the other and (2) both quantities have the same degree of smallness (*i.e.* they are equivalent to one another). These two cases do not, however, cover all the possible interrelationships between the order of decrease of two infinitesimal quantities which take part in the same process; on the contrary, the cases which we considered above are only the simpler instances and can be studied very easily. Generally speaking, the ratio y/x of two infinitesimal quantities can be much more complicated; for example, the quantity $|y/x|$ can, in the course of the process, become infinitesimal

and later infinitely large, and both phenomena can take place again and again no matter how far advanced the process. In this case we cannot ascribe to y (as compared to x) either a higher or a lower or even the same degree of smallness and we must admit that the quantities y and x *cannot be compared with one another* with respect to the rates of their decrease. From a logical point of view we should consider this case, where comparison is impossible, as the general case; however, in practice we meet more often one of the special cases which we have considered above.

All that has been said so far in this section about infinitesimal quantities can also refer, with appropriate corrections, to infinitely large quantities. Let us assume that the quantities x and y are infinitely large in a certain process. If the ratio x/y is infinitely large, then x is an infinitely large quantity of a higher order as compared to y and y is an infinitely large quantity of a lower order as compared to x . If $|x/y|$ from a certain moment onwards remains confined between two constant positive numbers, then x and y are infinitely large quantities of the same order; this will always be the case when there is a $\lim (x/y)$ in a given process which is other than zero; in particular, if $x/y \rightarrow 1$, the infinitely large quantities x and y are said to be equivalent and this is denoted by $x \sim y$. When evaluating limits we can replace infinitely large factors by arbitrary equivalent quantities in the same way as this was done with infinitesimal quantities.

As with infinitesimal quantities, it can be very useful to assess the order of infinitely large quantities not only qualitatively (higher, lower, equal) but also quantitatively. This can be done as follows. An arbitrary infinitesimal quantity, for example x , is chosen as the basis; all other infinitesimal quantities which are of the same order as x are said to be infinitesimal quantities of the *first order**; in particular, any infinitesimal quantity which is equivalent to x will be an infinitesimal quantity of the first order. To continue: the quantity x^2 and all quantities of an equal order are known as quantities of the *second order*. Generally speaking, any quantity of an order equal to that of x^α , where α is an arbitrary positive constant, is known as an infinitesimal quantity of *order* α . The rate of growth of infinitely large quantities is determined in a similar manner.

* Let us remember: y is of the same order as x if from a certain moment of the given process onwards we always have $a < |y/x| < b$, where a and b are positive constants.

Example 1. The problem solved at the end of the previous paragraph shows that if x is taken as the basis, then $1 - \cos x$ will be an infinitesimal quantity of the second order.

Example 2. Let

$$y = a_1 x^{n_1} + a_2 x^{n_2} + \dots + a_k x^{n_k},$$

where the constants a_1, a_2, \dots, a_k are other than zero and the positive numbers n_1, n_2, \dots, n_k are such that $n_1 < n_2 < \dots < n_k$. Thus (1) if x is the basic infinitesimal quantity, then y is an infinitesimal quantity of order n_1 ; (2) if x is an infinitely large quantity, then y is an infinitely large quantity of order n_k .

In conclusion of this paragraph we shall introduce another very convenient system of notation which is used more widely in contemporary mathematics and which we shall find very useful in future. Let y and x be two quantities which are involved in a certain process and let x always be positive, (or at least from a certain moment of the given process onwards). We then have: (1) if the ratio y/x is a bounded quantity in this process, then this is written down as follows:

$$y = O(x);$$

(2) if the ratio y/x is an infinitesimal quantity in the given process (*i.e.*, its limit is zero), then this is written as follows:

$$y = o(x).$$

It evidently follows from $y = o(x)$ that $y = O(x)$, but the converse is not true. It is self-evident that we assume in both these relations a definite process in which both quantities x and y participate and that, generally speaking, in any other process this will no longer be so.

Example 3. If x is an infinitesimal quantity, then

$$\begin{aligned} x^2 &= o(x), \\ 5x + 3x^2 &= O(x), \\ 2 \sin x &= O(x), \\ 1 - \cos x &= o(x). \end{aligned}$$

Example 4. If x is an infinitely large quantity, then

$$\begin{aligned} x &= o(x^2), \\ 5x + 3x^2 &= O(x^2). \end{aligned}$$

Example 5. For any arbitrary change in x , it follows from $y = o(x)$ that $x + y \sim x$; conversely, it follows from $x > 0$ and $x + y \sim x$ that $y = o(x)$.

Example 6. The fact that the quantity x is infinitesimal in the given process can be denoted as follows:

$$x = o(1) ;$$

and similarly, the relation

$$1 = o(x)$$

is equivalent to the statement that x is a positive infinitely large quantity in the given process and the relation

$$x = O(1)$$

is equivalent to the statement that the quantity x is bounded in the given process; we can thus see that the symbols O and o enable us, in certain cases to express very briefly the character of change of different quantities.

CHAPTER III

DEVELOPMENT OF THE ACCURATE THEORY OF LIMIT TRANSITION

§ 13. The mathematical definition of a process

Until now we assumed that all the quantities which we have considered above participated in a certain *process* (phenomenon) and tried to elucidate the character of their change in the course of this process. We were talking about different *moments* of the given process and differentiate between its *earlier and later moments*. This method of expression is picturesque, simple and convenient ; it helps the student to understand the origin of the main concepts of mathematical analysis (variable quantities, functions, limits) from his observations and studies of the outside world. However, from the point of view of *mathematical theory*, it is necessary to introduce greater accuracy into this form of expression; the concept of a process and its various moments, which we have used so far, was not defined mathematically and, when using this concept, we had in view no definite mathematical objectives apart from the convenience of the picturesque representation linked with our everyday experience. In order to become a fully valid science which can be subjected to mathematical investigation each process must be fully defined mathematically and be freed of ideas which are not defined in this way ; this form of description can serve as an abstract formal characteristic without which no mathematical theory can function.

When considering the real or abstract mathematical processes dealt with in earlier paragraphs, it can be readily seen that this mathematical characteristic, this formal structure of the process can be very different in different processes. However, there is one characteristic which all processes share in common and it is this characteristic which we must try to elucidate. This common characteristic is due to the fact that the different moments of any process are

always represented by a sequence of successive values of a certain variable which changes in which the given process is essentially concerned; it is therefore natural to call this quantity the *basic variable* of this process. Let us explain this by examples.

Example 1. In the course of the process described in example 3, § 7 (a geometrical progression, the n -th term of which is $1/2^n$) the n terms of the progression run successively through a series of natural numbers ($n = 1, 2, \dots$). By different “moments” of the process we mean different values of the number n , where the lower values of n correspond to the “earlier” moments and the higher values to the “later” moments of the given process. By the quantity which “participates” in this process we mean any function of n ; in particular one such function is the function $1/2^n$ which we have considered in this example. The “basic” variable of this process is the number n .

Example 2. In the example 1, § 7 (expansion of a gas at constant temperature) the volume v of the given mass of gas is the basic variable; in this process the value of v grows indefinitely. By different moments of the process we mean different values of v . As in the above example, the lower values of v refer to “earlier” moments and the higher values of v to “later” moments. In contrast to the above example where n acquires only integral values, the value of v increases continuously and in passing from one value to the next it runs through all intermediate values. Here again we take it that the quantity which “participates” in the given process can be any function of v ; in particular, one such function is $p = c/v$ which we have considered in example 1, § 7.

Example 3. Let us now consider a process which involves a continuous decrease in the positive number x ; we take it that this number is the basic variable of the given process. The “earlier” moments will be the greater and the “later” the smaller values of x . The quantity which participates in this process can be any function of x , for example, $1 + x + x^2$, $\cos x$, etc. In this case the basic variable behaves differently from that in either of the two processes considered above; it does not increase but decreases and tends towards zero. In example 4, § 10 we have analysed the behaviour of $\cos x$ which participates in a process of this kind.

Let us now draw some conclusions. We can see that from a mathematical point of view every process should be regarded as a series of successive values of a certain variable quantity, “basic” for the given process. The individual values of this variable represent

the moments of the given process, where a lower value corresponds to an earlier moment and a higher value to a later moment, or *vice versa*, (in other words, the basic variable either increases continuously in the given process or decreases continuously). The quantity which participates in the given process can be an arbitrary function of the basic variable.

Hence these are the common characteristics of the processes which we have studied so far. What, then, can be the differences between these processes from a mathematical point of view? If we were to disregard the real contents of these processes and were only to consider their mathematical structure, then, as we can see, the only difference lies in the behaviour of the basic variable. It is the character of the behaviour of this variable which influences the mathematical nature of the process and this, as we have seen above, can vary greatly. Apart from the three types of processes which we have considered above, many other cases are possible which have an even more complicated structure; thus, for example, we can imagine a process of "mixed" structure in which the basic variable changes either by jumps (as in the first example) or continuously (as in the other two examples); however, for the purpose of mathematical analysis, the structures which we have considered above are basically important, and therefore, we shall study them alone. We can, thus, regard every process as a series of successive values of a certain "basic" variable; the value of this variable can be expressed in terms of natural numbers (*i.e.*, it may change by jumps or increase continuously) or it changes continuously by running through intermediate values; in the latter case it can either increase continuously or it can decrease continuously; if, for example, it increases continuously, then it can either increase indefinitely or it can remain bounded; a continuously decreasing quantity can also behave in an analogous manner. In every case the character of change in the basic variable fully defines the mathematical type of process. As we know, these types can be very diverse; however, for the purpose of mathematical analysis, it is quite adequate to consider only a few simple types of processes mentioned above.

§ 14. The accurate concept of limits

In chapter 2 we agreed that a quantity y which participates in a given process tends towards the constant limit b if the difference $y-b$ is an infinitesimal quantity in this process; thus the concept of a limit is always given in terms of an infinitesimal quantity.

But what exactly do we mean by infinitesimal quantities? We have said that the difference $y-b$ is infinitesimal if *no matter how small the positive number ε be, the inequality $|y-b| < \varepsilon$ will always be satisfied from a certain moment onwards*. As we have already said in the previous paragraph, we can no longer be satisfied with this formulation which involves the concept of a process and its moments; for they are not defined accurately. However, we now know the accurate definition of a mathematical process; hence by replacing the indefinite terms of “process” and its “moments” used for picturesque description by the corresponding strictly accurate mathematical concept, we are fully capable of defining the concept of an infinitesimal quantity (and therefore also the concept of a limit) with absolute accuracy. The accurate definition of a limit, which we are trying to give in this paragraph, can be expressed in different ways for different types of processes; we shall therefore have to formulate it separately for every type of processes listed in the previous paragraph. This is in contrast to our earlier, not quite accurate definition of a process which could be formulated in the same way for every type of process; thus, by using this definition, we were able to define in chapter 2 the theory of limits for processes of all mathematical structures.

1. The limit of a sequence. Let us consider a process in which the basic variable n runs successively through a series of natural numbers ($n = 1, 2, \dots$). Any function of n can participate in the process, for example $1/2^n$, $n!$, the perimeter p_n of a regular n -sided polygon inscribed in a circle of unit radius, etc. Let a_n be one such function. In this process a_n runs successively through the sequence of values

$$a_1, a_2, \dots, a_n, \dots \quad (1)$$

Let us now try to explain accurately the statement that “the quantity a_n tends to a limit α in the given process”.

We know that the fundamental idea of this statement is as follows: no matter how small $\varepsilon > 0$ be, the inequality $|a_n - \alpha| < \varepsilon$ will be satisfied from a certain moment of the process onwards. But what is meant by “from a certain moment of the process onwards?” By the moments of our process we mean different values of the basic variable n —greater n corresponds to the later moment. Therefore the words “from a certain moment of the process onwards” mean more accurately “beginning with a certain value n_0 of the number n and for all its greater values”. The statement which we are trying to define accurately can therefore be formulated as follows: *the quantity*

a_n tends towards the limit α , if no matter how small the positive number ε , a natural number n_0 exists which is such that we have $|a_n - \alpha| < \varepsilon$ for any $n \geq n_0$. It is evident that this formulation is much more complicated than the one used previously, but it is quite free of concepts which are not fully defined ("process" and its "moments"). Therefore, this definition of a limit can now be used in the construction of a strict mathematical theory.

In processes of the type we are considering, it is more usual to speak not of a "function a_n " but of a "sequence of numbers" (1). In the event when $a_n \rightarrow \alpha$, the sequence (1) is said to be *convergent* and the number α is known as its limit. If a_n has no limit, then it is said that the sequence (1) is *divergent*.

The fact that the basic variable n increases indefinitely in the course of the process, is frequently denoted symbolically as follows: $n \rightarrow \infty$, or more accurately, $n \rightarrow +\infty$; it must be remembered that in this symbolic notation the arrow does not indicate a tendency towards a limit, as is usually the case, for an indefinitely increasing quantity can have no limit. The sentence "the sequence (1) tends to the limit α " (the exact meaning of which we now fully understand) can therefore, be expressed symbolically as follows:

$$\lim_{n \rightarrow \infty} a_n = \alpha,$$

or

$$a_n \rightarrow \alpha \quad (n \rightarrow \infty);$$

and both forms fully express the fact in which we are interested ($\lim a_n = \alpha$ or $a_n \rightarrow \alpha$) apart from indicating the nature of the process in which they take part ($n \rightarrow \infty$).

Example. Let us denote the sum of the first n terms of a geometrical progression

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2^n}, \dots,$$

by a_n ; we thus have

$$a_n = \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^n} = 1 - \frac{1}{2^n},$$

and the sequence (1) becomes

$$1 - \frac{1}{2}, 1 - \frac{1}{2^2}, 1 - \frac{1}{2^3}, \dots, 1 - \frac{1}{2^n}, \dots$$

We evidently have $a_n \rightarrow 1$ ($n \rightarrow \infty$). Thus owing to the fact that

$$|a_n - 1| = \frac{1}{2^n} \quad (n = 1, 2, \dots),$$

and no matter how small $\varepsilon > 0$, we can choose n_0 so great that $1/2^{n_0} < \varepsilon$, so that for every $n \geq n_0$:

$$|a_n - 1| < \varepsilon.$$

2. The one-sided limit of a function. We shall now consider the second basic type of process, *i. e.* a process in which the basic variable x varies continuously, *i. e.* it runs through all intermediate values; in doing so it can either increase indefinitely or decrease indefinitely and thus remain limited, or it can become an indefinitely large quantity, *i. e.* its absolute value can grow indefinitely. Each case must be considered on its own merits; however, all cases have many common characteristics which enable us to treat them more briefly. In every case the quantity y which participates in a process is assumed to be an arbitrary function $y = f(x)$ of the basic variable x and we are here trying to explain clearly the exact meaning of the statement "in the given process the quantity y tends to the limit b ".

Let us first consider the case when the basic variable x grows indefinitely and is a positive number ($x \rightarrow +\infty$)*.

This case is very close to the previous case and the only difference is due to the fact that here x runs through all intermediate values in the process of its growth, whereas in the other example n could assume only integral values. As before the words "from a certain moment onwards" mean "beginning with a certain value A of x and for all its greater values". The exact meaning of the statement

$$\lim_{x \rightarrow \infty} y = b \quad \text{or} \quad y \rightarrow b \quad (x \rightarrow \infty)$$

here is as follows: *no matter how small the positive number ε is, there exists a positive number A such that $|y - b| < \varepsilon$ for any $x \geq A$.*

In cases where the basic variable x is a *negative* infinitely large quantity ($x \rightarrow -\infty$), *i. e.* although it is a negative quantity its

* The notation $x \rightarrow \infty$ or $x \rightarrow +\infty$ is used to denote the indefinite growth of x , both when this growth takes place by jumps and when it is continuous; for this reason the actual type of variation should in each case be indicated in the text.

absolute value grows indefinitely, the relationship $y \rightarrow b$ is obviously defined in the same way: the relationship

$$\lim_{x \rightarrow -\infty} y = b \quad \text{or} \quad y \rightarrow b \ (x \rightarrow -\infty)$$

means that *no matter how small $\varepsilon > 0$ may be, there exists a positive number A such that $|y - b| < \varepsilon$ for any $x \leq -A$.*

Let us now consider the case when the basic variable changes continuously (*i.e.* increases or decreases continuously) and remains, at the same time, a limited quantity. We shall learn in chapter IV that in this event x tends to a certain limit a . If x grows indefinitely, then it approaches a from the side of the lower values ("from the left") and this is usually denoted by: $x \rightarrow a - 0$. If x decreases continuously then it always remains greater than the number a and approaches it from the side of the greater values ("from the right"); this is denoted as: $x \rightarrow a + 0$. To begin with let us consider the first case ($x < a$, $x \rightarrow a - 0$). The words "from a certain moment of the process onwards" evidently mean here "beginning with a certain value $a - \delta < a$ of x , and for all values closer to a (and, of course, smaller than a); more briefly we can say: "for all values of x which satisfy the inequalities $a - \delta \leq x < a$ ". The exact meaning of the statement

$$\lim_{x \rightarrow a-0} y = b \quad \text{or} \quad y \rightarrow b \ (x \rightarrow a - 0)$$

is as follows: *no matter how small the number $\varepsilon > 0$, there exists a positive number δ such that $|y - b| < \varepsilon$ for any value of x which satisfies the inequalities $a - \delta \leq x < a$.*

The exact meaning of the statement given below is determined in a similar way

$$\lim_{x \rightarrow a+0} y = b \quad \text{or} \quad y \rightarrow b \ (x \rightarrow a + 0)$$

and is formulated in exactly the same way except that the inequality $a - \delta \leq x < a$ must be replaced by $a < x \leq a + \delta$.

We have thus established the exact meaning of the concept of limit for all processes of the basic types which are used in mathematical analysis. Let us emphasize once again the fact that all arguments and results obtained in chapter II which were the same for all types of processes, are, of course, also valid for our new and more exact definition of limit transitions; the new definition does not

contradict in any way our old definition with which it is compatible—it only provides more exact specifications for cases of different kinds.

Let us now make one more remark. Let us assume that the quantity y which participates in a certain process, does not tend towards a limit, but continues to grow indefinitely. Let us assume that we are dealing with a process of the type $x \rightarrow a - 0$; we can, therefore, write

$$y \rightarrow +\infty \quad (x \rightarrow a - 0).$$

What is the exact meaning of this statement? From all that was said above, we are able to answer this question without difficulty: *no matter how small $A > 0$, there can be found a $\delta > 0$ so that we have $y > A$ for all values of x which satisfies the inequalities $a - \delta \leq x < a$.*

By using this as an example, the reader will be able to find for himself the exact meaning of the relationships $y \rightarrow +\infty$ and $y \rightarrow -\infty$ in any process of the type considered above. He will find this to be an excellent exercise.*

§ 15. The Development of the Concept of Limit Transitions

The two types of limit transitions (the limit of a sequence and the one-sided limit of a function), which we considered in great detail in the previous paragraph, are of basic importance in mathematical analysis, for, all other more complicated types of processes which we shall encounter in future can be broken down to those cases. However, to make this reduction possible in every case, we must now develop somewhat the concept of a quantity which tends towards a limit in a given process.

Let us begin by considering a simple example which will show us the necessity and the course of this development. Let us assume that the process in which we are interested involves infinite decrease of the perimeter p ("basic" variable) of a certain rectangle where the form of this rectangle can change in the course of the process in any way we please. Owing to the fact that in a rectangle of perimeter p , each side is smaller than $p/2$, the area s of the rectangle of perimeter p will always be smaller than $p^2/4$. When $p \rightarrow 0$, we

* Cf. problems 349-352, section 1 of the 'problem book' by B. P. Demidovich mentioned in the preface. The numbers of the problems as they appear in the second edition, can be found on p. 622.

evidently have $p^2/4 \rightarrow 0$; therefore, the area s in our process (i.e. when $p \rightarrow +0$), is an infinitely small quantity, and we can write

$$s \rightarrow 0 \quad (p \rightarrow +0).$$

The exact meaning of this statement is determined in the usual way: *no matter how small $\varepsilon > 0$, there exists a $\delta > 0$ such that the area of any rectangle, the perimeter of which is smaller than δ , will be smaller than ε .*

This example differs basically from all the examples which we considered so far. This difference is due to the fact that for a given perimeter p , the area s of a rectangle can have an infinite number of different values, so that s is not a function of p . Owing to the fact that we have taken p as the basic variable in our process, and also because so far we assumed that the quantity participating in a given process is a function of the basic variable, we cannot, strictly speaking, consider s to be a quantity which participates in our process; it is even less possible to speak of its tendency towards a limit. Here we are dealing with a quantity whose value at every moment of the process (i.e., for every value of p) remains indefinite. At the same time it is still true to say that provided the perimeter p of the rectangle is chosen sufficiently small, the area s of this rectangle, *no matter what infinite number of possible values it may assume*, will be as small as we please. More exactly: *no matter how small $\varepsilon > 0$, there is a $\delta > 0$, such that for any rectangle with perimeter $p < \delta$, we shall have $s < \varepsilon$, where s is any possible value of the area of a rectangle with a perimeter p .*

Hence the accurate meaning of the relation

$$s \rightarrow 0 \quad (p \rightarrow +0)$$

which is generally accepted, remains valid for our example, in spite of the fact that s is not a function of p . It is, therefore, possible to apply to examples of this kind all propositions of general theory which were stated in chapter II. It is now only necessary to develop our mathematical interpretation of the phrase "a quantity participating in a given process" and to determine the concept of the tendency towards a limit as applied to this extended class of quantities. The above example shows quite clearly that this must now be done. To begin with, from now on we shall understand by a quantity which participates in a process, *any quantity y with regard to which it is known what values it can assume for any given value of the basic variable x (i.e., at any given moment of the process)*; it is thus evident that our former agreement that y must always be a function of x is a particular case

in the wider definition which we now accept; we arrive at this particular case by assuming that the set of values which y can take for the given value of x , is always a single number. Let us assume that our process is described by the relationship $x \rightarrow a + 0$. In this case the exact meaning of the statement that in the given process $\lim y = b$ (where y is a quantity participating in this process as given by our wider definition) involves the following: *no matter how small $\varepsilon > 0$, there is a $\delta > 0$, such that for any value of x confined between a and $a + \delta$, and for any y possible for the given value of x , we have :*

$$|y - b| < \varepsilon.$$

If this requirement is satisfied, we can write :

$$y \rightarrow b \quad (x \rightarrow a + 0).$$

Two-sided limit of a function. Let us now consider an important example of the application of the wider definition of the concept of limit transition as given above. This example will show that our wider definition is already useful during our first attempt in mathematical analysis.

Let us assume that $y = f(x)$ is a function of x and let the value of y get as close as we please to a number b when the value of x is sufficiently close to the number a (and is, at the same time, other than a). By now we are well-acquainted with the exact meaning of statements of this kind: *no matter how small $\varepsilon > 0$, there is a $\delta > 0$, such that for every $0 < |x - a| \leq \delta$, we have $|y - b| < \varepsilon$.* Symbolically this can be written as follows

$$y \rightarrow b \quad (|x - a| \rightarrow + 0). \quad (1)$$

According to our system of notation this symbol means that the quantity $|x - a|$ is the basic variable in the process under consideration and that y tends to the limit b in this process. But every given value $|x - a| = \alpha$ of the basic variable $|x - a|$ corresponds to two different values of x : $x = a + \alpha$, $x = a - \alpha$, and therefore, there are, generally speaking, two different values of y : $y = f(a + \alpha)$ and $y = f(a - \alpha)$. Hence for any value α of the basic variable, the quantity y can assume two different values and therefore, it is not a function of a single basic variable. Nevertheless, our wider definition of limit transition enables us to write the relationship (1) and to maintain that y tends towards b as its limit, when $|x - a| \rightarrow + 0$.

By the way, the process $|x - a| \rightarrow +0$ is usually written in the form $x \rightarrow a$, so that instead of the relationship (1) we can write

$$y \rightarrow b \quad (x \rightarrow a). \quad (2)$$

The notation $x \rightarrow a$, in contrast to the former notations $x \rightarrow a-0$ and $x \rightarrow a+0$, shows that in this case, x approaches the number a , but that it must not necessarily increase or decrease: it can change the direction of its transition and, in particular, become greater or smaller than a . Therefore, the limit of y when $x \rightarrow a$, which we just described, is known as *the two-sided limit of a function*.

Let us remember once again that the exact meaning of the limit transition (2) involves the following: *no matter how small $\varepsilon > 0$, there is a $\delta > 0$, such that for any value of x for which $0 < |x - a| \leq \delta$, we have $|y - b| < \varepsilon$.*

Let us now make one more important remark: *in order that the number b should be the two-sided limit of y when $x \rightarrow a$, it is necessary and sufficient that the one-sided limits of y , viz. $\lim_{x \rightarrow a+0} y$ and $\lim_{x \rightarrow a-0} y$, should exist and be equal to b .* In fact let us assume that $\varepsilon > 0$ is given arbitrarily. If

$$\lim_{x \rightarrow a} y = b,$$

then provided we have a sufficiently small $\delta > 0$, it follows from $0 < |x - a| \leq \delta$ that $|y - b| < \varepsilon$. But if $a < x \leq a + \delta$, then even more so, $|x - a| \leq \delta$, and therefore, we also have $|y - b| < \varepsilon$. Thus when $y \rightarrow b$ ($x \rightarrow a+0$), it can be shown similarly that this is also true when $y \rightarrow b$ ($x \rightarrow a-0$). Let us now assume, conversely, that we are given $y \rightarrow b$, when $x \rightarrow a+0$, and when $x \rightarrow a-0$. In that case, *no matter how small $\varepsilon > 0$ we can find a δ_1 which is such that $|y - b| < \varepsilon$, when $a < x \leq a + \delta_1$, and such a δ_2 , that $|y - b| < \varepsilon$ when $a - \delta_2 \leq x < a$; if we denote by δ the smaller of the numbers δ_1 and δ_2 , then $a - \delta \leq x \leq a + \delta$ ($x \neq a$), and we have $|y - b| < \varepsilon$; this shows that $y \rightarrow b$ ($x \rightarrow a$), which had to be proved.*

We can thus see that the process of two-sided approach of the variable x to the limit a simply involves the process of the one-sided approaches $x \rightarrow a+0$ and $x \rightarrow a-0$. This is the first example illustrating what was said in the note in § 14, i.e. that different types of analyses can be reduced to the study of the two basic types of processes.

CHAPTER IV

REAL NUMBERS

§ 16. Necessity of Producing a General Theory of Real Numbers

One characteristic of a variable quantity arises from the fact that in the course of a process it assumes different values. Each of these values is expressed in terms of some number. If, for example, temperature of air rises from 5 to 10°C , we naturally assume that in the course of this process it runs gradually through *all* the numbers from 5 to 10 . But what exactly do we mean by “all the numbers”? It is evident that these numbers are not restricted to integers alone, for obviously there are moments when the temperature is equal to 6.5°C . Do we, perhaps, mean “all integers and all fractions”?

The set of all integral and fractional numbers (positive, negative and zero) forms the so-called set of *rational numbers*; these numbers and the operations which can be performed with them are studied in detail, in arithmetic and algebra. The question whether these numbers are sufficient for measuring all the quantities which we are likely to meet in the study of the world around us is of great importance, both in mathematics as well as in the accurate study of nature. In ancient Greece (probably in the Pythagoras school) a remarkable discovery was made, *viz.* that certain simple geometrical constructions lead easily and irrevocably to quantities which cannot be measured with the help of rational numbers. A simple case of this type is very well-known: if each side adjacent to the right angle in a right-angled triangle is of unit length, then according to the theorem of Pythagoras, the hypotenuse of this triangle should be such that its square is equal to 2 . But it is easy to show that there is no rational number whose square

is equal to 2^* . Therefore, if we want to restrict ourselves to rational numbers, we must admit that the hypotenuse of the triangle in question has no length; obviously we cannot arrive at this conclusion, for geometry cannot be based on this absurdity. Circumstances in the outside world thus make it impossible for us to restrict ourselves to the set of rational numbers alone; it is therefore necessary to add a new type of numbers which we shall call *irrational*. One such number is $\sqrt{2}$, the square of which, by definition, is equal to 2. However, it must be remembered that the introduction of a new number is an easy matter which does not by itself have any significance; if we wish to make this newly-introduced number a fully valid number of the family of numbers, we must, to begin with, define its position in this family, *i.e.* we must determine which rational numbers are smaller and which greater than $\sqrt{2}$. In the second place, we must define all operations to which this new number can be subjected, (for we do not know, for example, what is meant by $\sqrt{2} + 1$, $3\sqrt{2}$, $1/\sqrt{2}$), and we must prove that these operations are subject to the same laws which govern operations with rational numbers (for example, we must show that $\sqrt{2} + 1 = 1 + \sqrt{2}$). All this can be done, though it would necessitate considerable effort; however, the object we have in mind fully justifies this procedure. But let us assume that we already did all this. Sooner or later we shall meet another physical or geometrical problem which will make necessary to introduce another new number, the square of which is equal to 3 or 5, etc. It would thus no longer be possible to repeat in each case the same chain of arguments which we used for making $\sqrt{2}$ a fully valid number. Let us now assume that we have found a way which would enable us to use a single method for introducing square roots of all natural numbers into the family of numbers (this is, no doubt, possible). The possibilities of applications are hereby not exhausted. If we are trying to find the length of the side of a cube, whose volume is equal to 2 m^3 , we must introduce the number $\sqrt[3]{2}$. And even if we do introduce roots of any degree of any rational number into the family of numbers, this will not be sufficient. On one hand, the required number is frequently defined as a root of a given equation; on the other hand, we know practically

* If we have $(p/q)^2 = 2$, then we find that $p^2 = 2q^2$; let $p = 2^r p'$, $q = 2^s q'$, where p' and q' are no longer even numbers. Then we have

$$p^2 = 2^{2r} p'^2, \quad 2q^2 = 2^{2s+1} q'^2,$$

and the equality $p^2 = 2q^2$ leads to a contradiction, for its left side contains 2 with an even index and the right side contains 2 with an odd index.

that there must be one such root, whereas theory shows that there is no such root among all possible rational and irrational numbers; and again we find it necessary to introduce a new number which we simply define as the root of our equation. Here again we must repeat the same argument which we used above in connection with the number $\sqrt{2}$. In practice, even the simplest geometrical problems may lead to difficulties of this kind. This is particularly shown by the following example in which we try to find the area of a circle of unit radius. We know that the area of a circle is defined as the limit of the areas of all inscribed (or circumscribed) regular polygons when the number of sides of these polygons increases indefinitely. We know from this visual representation and practice that a circle has an area; in reality, can we reconcile ourselves to the fact that such a simple figure as a circle has no area at all? At the same time mathematics tells us that there is no such limit among all the numbers so far handled, including the roots of all algebraic equations. Hence we have no alternative but to introduce a completely new number for measuring the area of our circle and repeat once again the chain of arguments mentioned above, in order to make this new number a fully valid member of our extended family of numbers. This new number is no other than the well-known number π .

The above examples clearly show that the procedure is unscientific and unpracticable, if, in order to solve a problem, for whose solution the existing numbers are insufficient, we find it necessary to introduce new numbers, define their position among the existing numbers, find and investigate the operations which can be performed with them, etc.,—in other words—to do all that is necessary to make them fully valid members of the family of numbers. It is thus quite clear that a *general theory of irrational numbers* must be produced; we must find one general principle of origin of irrational numbers (the numbers studied so far are particular cases) which would include all the historically known examples of this kind, and thus guarantee that it will no longer be necessary to introduce further new irrational numbers. For numbers originated by this general principle, it will be necessary to repeat all arguments in general form, but this will in future enable us to operate with them in the same way as we do with rational numbers in elementary arithmetic and algebra. This is the only scientific approach to the problem in question.

All this work is not a part of mathematical analysis—a science which deals with changes in quantities—but is part of the theory of

numbers; however, until this problem is solved, mathematical analysis can have no stable basis; in fact, as we have already said at the beginning of this paragraph, the values of all variable quantities are expressed in terms of numbers; therefore, we cannot even begin the study of variable quantities without knowing the numbers which modern mathematics has at its disposal and the properties of this set of numbers. A short outline of the modern theory of this set of numbers is given in the next few paragraphs of this chapter.

§ 17. Construction of a Continuum

1. When we evaluate $\sqrt{2}$ with the help of conventional methods we obtain the following sequence of approximations for this number :

$$a_0 = 1; \quad a_1 = 1.4; \quad a_2 = 1.41; \quad a_3 = 1.414; \dots$$

Each one of these values is a rational number (a finite decimal fraction) and each number is greater than the preceding number (or, at least, is equal to it). The squares of these numbers tend to 2. *

$$a_n^2 \rightarrow 2 \quad (n \rightarrow \infty).$$

However, the numbers a_n cannot tend to a *rational* limit: if such a limit r exists, then $a_n \rightarrow r$ would imply $a_n^2 \rightarrow r^2$, and since $a_n^2 \rightarrow 2$, we would have $r^2 = 2$; but this would mean that a rational number r exists such that its square is equal to 2, which, as we know, is incorrect.

We thus have the sequence

$$a_0, a_1, a_2, \dots, a_n, \dots \quad (1)$$

of well-known numbers; this is an *increasing* sequence, *i.e.* we always have $a_{n+1} \geq a_n$; at the same time this sequence has no rational limit. Wherever we are introducing the new (irrational) number

* In fact, each one of the numbers a_n is such that if we take its last decimal place by one unit greater, we obtain a number whose square is > 2 ; hence

$$a_n^2 < 2 < \left(a_n + \frac{1}{10^n}\right)^2,$$

and therefore,

$$0 < 2 - a_n^2 < \left(a_n + \frac{1}{10^n}\right)^2 - a_n^2 = \frac{2a_n}{10^n} + \left(\frac{1}{10^n}\right)^2 \rightarrow 0 \quad (n \rightarrow \infty).$$

$\sqrt{2}$ whose square is by definition equal to 2, we are filling in, as it were, a gap existing between rational numbers: our number must fill this gap in the set of rational numbers and be defined as limit of the increasing sequence (1).

The situation created by the introduction of the irrational number π is very similar to the one which we have just described. Let us assume that the area of a regular n -gon inscribed in a circle of unit radius is equal to s_n ; in this case the numbers

$$s_3, s_4, s_5, \dots, s_n, \dots \quad (2)$$

form an increasing sequence and the number π is defined geometrically as limit of this sequence. Here the position is somewhat complicated by the fact that the area s_n is, generally speaking, expressed in terms of irrational numbers; however, these numbers are among the simple irrational numbers and can easily be expressed in terms of roots of natural numbers; we can, therefore, assume that the area s_n is expressed by a well-known number. It now appears that the sequence (2) has no limit either among rational numbers or even among the numbers of the wider class in terms of which the area s_n is expressed. Thus by re-introducing our new number π we are filling, as it were, a gap existing in the set of all the numbers we have met so far, and this number is the limit of the increasing sequence (2), i.e. it is a limit which did not exist among the numbers we have known so far.

Let us assume now that we are given an arbitrary increasing sequence

$$r_1, r_2, \dots, r_n, \dots \quad (r_{n+1} \leq r_n) \quad (3)$$

of *rational* numbers. To begin with, we must distinguish two cases: the number r_n can grow indefinitely as n increases; or a positive number C can exist such that $r_n < C$ for any n . In the first case r_n is an infinitely large quantity when $n \rightarrow \infty$, and therefore, it cannot tend to a limit. We shall, therefore, concentrate on the second case, remembering, that for the moment we only have rational numbers at our disposal. In the case under consideration the sequence (3) is *bounded*; it may happen, however, that it has a rational limit r ; thus the sequence

$$r_1 = 1 - \frac{1}{1}, \quad r_2 = 1 - \frac{1}{2}, \dots, \quad r_n = 1 - \frac{1}{n}, \dots$$

is an increasing bounded sequence which tends to unity as its limit :

$$r_n = 1 - \frac{1}{n} \rightarrow 1 \quad (n \rightarrow \infty).$$

It may also happen that the bounded increasing sequence has no rational limit; thus, for example, the sequence (1) of approximated values of $\sqrt{2}$ is evidently an increasing bounded sequence (all $a_n < 2$), but at the same time we have seen that it has no limit.

Let us now agree (in the same way as we did when introducing the irrational number $\sqrt{2}$) that *every time when we deal with a bounded sequence (3) of rational numbers, for which there is no rational limit, we shall take a new irrational number as its limit.** We have thus established a general *principle of origin* of irrational numbers. Having made this agreement, we have also defined the whole set of irrational numbers. We shall see later that the set satisfying this definition has, in fact, taken some final form; in future we shall not introduce other new numbers apart from those defined by our agreement.

2. Example. Let us assume that $a_n = (1 + 1/n)^n$, ($n = 1, 2, \dots$), so that all the numbers a_n are rational ($a_1 = 2$, $a_2 = 9/4$, $a_3 = 64/27$, etc.). We will show that the sequence of numbers a_n is an increasing bounded sequence and that has an upper limit. According to the binomial formula we have :

$$\begin{aligned} a_n &= \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^n C_n^k \left(\frac{1}{n}\right)^k \\ &= 1 + \sum_{k=1}^n \frac{n(n-1) \dots (n-k+1)}{k!} \cdot \frac{1}{n^k} \\ &= 1 + \sum_{k=1}^n \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} \\ &= 1 + \sum_{k=1}^n \frac{1}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \\ &\quad \dots \left(1 - \frac{k-1}{n}\right). \quad (4) \end{aligned}$$

* We shall show at the end of this paragraph that this number does, in fact, satisfy the definition of a limit.

Similarly,

$$\begin{aligned}
 a_{n+1} &= \left(1 + \frac{1}{n+1}\right)^{n+1} \\
 &= 1 + \sum_{k=1}^{n+1} \frac{1}{k!} \left(1 - \frac{1}{n+1}\right) \left(1 - \frac{2}{n+1}\right) \cdots \\
 &\quad \cdots \left(1 - \frac{k-1}{n+1}\right). \quad (5)
 \end{aligned}$$

A comparison of the right-hand sides of the formulae (4) and (5) shows that in the sum (5) each term is greater than its corresponding term in the sum in formula (4), since the replacement of n by $n+1$ causes an increase in each small bracket in formula (4); moreover, there is an additional term in formula (5) corresponding to $k = n+1$ which is absent in formula (4). Therefore

$$a_{n+1} > a_n \quad (n = 1, 2, \dots),$$

i.e. the sequence of numbers a_n is increasing. It also follows from formula (4) that for any n

$$a_n \leq 1 + \sum_{k=1}^n \frac{1}{k!},$$

and owing to the fact that $k! \geq 2^{k-1}$ for $k \geq 1$, we have

$$a_n \leq 1 + \sum_{k=1}^n \frac{1}{2^{k-1}} = 1 + 2 \sum_{k=1}^n \frac{1}{2^k} < 1 + 2 \sum_{k=1}^{\infty} \frac{1}{2^k} = 3$$

and this shows that the sequence of numbers a_n has an *upper limit*.

In accordance with the accepted principle of origin we should therefore, assume that this sequence has a limit $\lim_{n \rightarrow \infty} a_n = e$ (rational or irrational). Further analysis, which we are unable to give here, shows that the number e is an irrational number.* We shall later see that this number e , like π , is one of the most important numbers in mathematical analysis; we shall meet it again in many chapters of our course. The first few decimal places of this number are as follows: $e = 2.71828 \dots$

* This means that the sequence of numbers a_n has no rational limit.

It is obvious that the above argument is only the first step to produce a general theory of irrational numbers. All rational numbers and all irrational numbers which we introduced on the basis of the principle of origin are *real numbers*. The set of all numbers is called *continuum*. The continuum represents the set of "values" which can be assumed by a "continuously" varying quantity. The basic aims of the theory of continuum are as follows : (1) to introduce "order" among the set of real numbers, *i.e.* to determine under what conditions one real number is greater than, equal to, or smaller than the other; (2) to determine all algebraic operations which can be performed with all real numbers and (3) to establish the laws which govern these operations. All these problems are satisfactorily solved by modern mathematical theory—all operations within our wider region of numbers are subjected to exactly the same laws as operations with rational numbers. Moreover, the number of operations which can be performed becomes wider : thus when dealing with real numbers, we can, for example, extract roots of any natural degree of any number (except for roots of an even degree of negative numbers, for they are no longer real, but imaginary numbers, and we are omitting them here). Within the scope of this course of mathematical analysis, we are unable to give sufficient attention to the development of this theory and we must, therefore, accept its conclusions as a ready basis for our further investigations. We shall thus restrict ourselves to a few brief remarks on this problem. The reader who is not interested in the theory of continuum can omit the points 3, 4 and 5 of this paragraph and pass on directly to point 6.

3. Let us assume that the increasing sequence $r_1, r_2, \dots, r_n, \dots$ defines a real number α . If all the numbers r_n , from a certain number k onwards, are equal to one another, *i.e.* $r_k = r_{k+1} = r_{k+2} = \dots = r$, then evidently $\alpha = r$, and α is a rational number. This sequence r_n is called *stationary* sequence; it is obvious that if α is an irrational number, the sequence determining this number is never stationary; however, if α is rational, the sequence r_n can be stationary [$r_n = \alpha, n = 1, 2, \dots$], or non-stationary [$r_n = \alpha - 1/n, n = 1, 2, \dots$]. This shows that in the construction of continuum, we could restrict ourselves to the consideration of non-stationary, increasing sequences of rational numbers.

Let us assume that we are given two non-stationary increasing bounded sequences of rational numbers

$$r_1, r_2, \dots, r_n, \dots, \quad (r)$$

$$s_1, s_2, \dots, s_n, \dots \quad (s)$$

As we know, each sequence originates a real number which can be rational or irrational. Let us assume that α is this number [for the sequence (r)] and β [for the sequence (s)]. We must now solve the problem, which of the three possible relations $\alpha < \beta$, $\alpha > \beta$, $\alpha = \beta$ applies in this case.

Let us agree to say that the sequence (s) is a *major* (exceeding) sequence as compared to the sequence (r) , if for every number r_n of the sequence (r) , a number s_m of the sequence (s) can be found such that $s_m \geq r_n$ (meaning of this inequality is clear, since the numbers r_n and s_m are rational). We can have four different cases :

- (1) (s) is major as compared to (r) and (r) is major as compared to (s) ;
- (2) (s) is major as compared to (r) but (r) is not major as compared to (s) ;
- (3) (r) is major as compared to (s) but (s) is not major as compared to (r) ;
- (4) (s) is not major as compared to (r) and (r) is not major as compared to (s) .

It can readily be seen that the fourth case is impossible. In fact, if (s) is not major as compared to (r) , then a number r_n can be found such that $s_m < r_n$ for any m ; but it is obvious that in this case (r) is major as compared to (s) . We, therefore, only have to consider the first three cases. In case (1) we are assuming that $\alpha = \beta$, in case (2) that $\alpha < \beta$ and in case (3) that $\alpha > \beta$. These assumptions uniquely define which of the three relations applies to any pair of real numbers. It can readily be shown that in the event when both numbers α and β are rational and (r) and (s) are strictly increasing sequences, the above concepts of equality and inequality would, as expected, coincide with the conventional concepts.

We must now find out whether the definitions of equality and inequality of real numbers, as given above, possess the same properties as those for rational numbers. Let us consider, for example, the *transitive* property, which is due to the fact that $\alpha \leq p$ and $p \leq \gamma$ implies $\alpha \leq \gamma$. This is a well-known property of rational numbers; but for real numbers it must be proved on the basis of our definition of equality and inequality. This can be done quite easily: to begin with, we must establish the transitive property of majority, *i.e.* if (s) is major as compared to (r) and (t) is major as compared to (s) then (t) is major as compared to (r) .

4. After establishing all necessary properties of equality and inequality, the theory of continuum tries to establish the *operations* to be performed with real numbers, *e.g.* how to determine the sum $\alpha + \beta$ of two real numbers. Let us assume that α is defined by the sequence (r) and β by the sequence (s) ; then

$$r_1 + s_1, r_2 + s_2, \dots, r_n + s_n, \dots, \quad (t)$$

is evidently an increasing bounded sequence of rational numbers; the real number γ thus defined is naturally the sum $\alpha + \beta$ of the numbers α and β . It can readily be shown that when α and β are rational, this definition of a sum coincides with the conventional definition. It can also be shown that this definition of addition conserves all rules of operations as they apply to rational numbers; thus, for example, the *interchange of terms* in addition (*i.e.* the rule $\alpha + \beta = \beta + \alpha$) follows directly from the definition, because an interchange of positions of the sequences (r) and (s) does not alter the sequence (t) .

Other operations with real numbers are determined in a similar manner, and the properties of these operations can be shown to be the same as those of the corresponding operations with rational numbers. We shall not consider subtraction, multiplication and division of real numbers, raising these numbers to integral positive powers or extraction of roots of integral positive powers. We shall only consider the definition of the expression a^x , where $a > 0$ and x is any real number (the definition of an exponential function). Let us assume that $a > 1$; then the rational a^x is also defined for any x , and is a rational function of x ; in fact, if $r = p/q$ and $r' = p'/q$ are rational numbers and if $r < r'$ ($p < p'$),

then $a^{\frac{1}{q}} > 1$ and therefore $a^r = \left(a^{\frac{1}{q}}\right)^p < \left(a^{\frac{1}{q}}\right)^{p'} = a^{r'}$.

Let us now assume that the real number α is defined by the increasing bounded sequence (r) of rational numbers. In this case the sequence

$$a^{r_1}, a^{r_2}, \dots, a^{r_n}, \dots,$$

is evidently bounded, and it follows from the above proof that it is also an increasing sequence. Hence it defines some real number which is evidently the number a^α . In this way the exponential function a^x is defined for any real x ; at the same time we also establish the fact that this is an increasing function (if $a > 1$), and a decreasing function (if $a < 1$). A logarithmic function is defined in a similar way.

It follows from these definitions that the known properties of the rational values of arguments of these functions also apply to all real values; thus we have in all cases $a^{x+y} = a^x a^y$, $\log_a(xy) = \log_a x + \log_a y$, etc.

We are unable to pay greater attention to these problems within the scope of this course.

5. There is, however, one more point which must be clarified. In the above wording of the principle of origin of real numbers, we said that the number α has originated from the increasing bounded sequence of rational numbers

$$r_1, r_2, \dots, r_n, \dots \quad (t)$$

and is assumed to be *limit* of this sequence. To convert this statement into a real tool of investigation, we must *prove* it; having learnt the arithmetic of real numbers, we are now, in principle, able to do it. It is evidently necessary to prove that no matter how small $\varepsilon > 0$, we shall have for all sufficiently large n :

$$\alpha - r_n < \varepsilon.$$

To begin with, let us prove the following auxiliary result on sequences of rational numbers.

Lemma. *Let (r) be an increasing bounded sequence of rational numbers. In that case, no matter how small $\varepsilon > 0$, there is an index n_0 , such that $n > n_0$ for $m > n_0$, and we always have $r_n - r_m < \varepsilon$.*

Proof. Let us assume that the statement expressed by the lemma is incorrect, i. e. that there is an $\varepsilon > 0$ such that the inequality

$$r_n - r_m \geq \varepsilon$$

is satisfied for all values of n and m which can be as large as we please. In this case, no matter how large the natural number k , there are k pairs of indices (m_i, n_i) ($1 \leq i \leq k$) such that

$$m_1 < n_1 < m_2 < n_2 < \dots < m_k < n_k$$

and

$$r_{n_i} - r_{m_i} \geq \varepsilon \quad (1 \leq i \leq k);$$

but in this case

$$\begin{aligned} r_{n_k} - r_{m_1} &= (r_{n_k} - r_{m_k}) + (r_{m_k} - r_{n_{k-1}}) + (r_{n_{k-1}} - r_{m_{k-1}}) \\ &\quad + (r_{m_{k-1}} - r_{n_{k-2}}) + \dots + (r_{n_1} - r_{m_1}) \geq k\varepsilon, \end{aligned}$$

for the first, third, fifth, etc. brackets are not smaller than ε and the second, fourth, etc. brackets are positive. Hence

$$r_{n_k} \geq k\varepsilon + r_{m_1}.$$

Owing to the fact that k can be as large as we please, the sequence (r) will contain terms which are as large as we please, and this contradicts the fact that it is a bounded sequence. This proves our theorem.

Let us now assume that ε is an arbitrary *rational* number; it follows from the above lemma that $r_n - r_k < \varepsilon$ for a sufficiently large k and for any n ; therefore, the sequence

$$\varepsilon, \varepsilon, \dots, \varepsilon, \dots$$

is a major sequence as compared with

$$r_1 - r_k, r_2 - r_k, \dots, r_n - r_k, \dots,$$

which evidently gives the real number $\alpha - r_k$; owing to the fact that the sequence (ε) gives the number ε , we have by virtue of the definition of inequality of real numbers

$$\alpha - r_k < \varepsilon.$$

for sufficiently large k .

This proves the proposition for a *rational* ε ; but owing to the fact that for any real $\varepsilon > 0$, we can find $\varepsilon' > 0$, smaller than ε , we would prove the proposition for any real $\varepsilon > 0$.

We must also draw attention to the fact that the principle of origin of irrational numbers accepted by us is by no means the only possible method; in the second half of the last century, when the necessity of producing a general theory of real numbers became apparent, several such theories were advanced almost simultaneously and each theory had its own principle of origin; it later became obvious that all these theories are basically equivalent so that the choice of theory should be governed not so much by principle as by the convenience of the method of treatment and its applications.

6. The wider set of numbers which we are now studying, is, as we know, by no means the first in the historical development of numbers. To begin with, we learn about the natural numbers in arithmetic, to which subsequently zero, negative and fractional numbers were added. Thus as a result of successive additions, the set of rational numbers is obtained. Our principle of origin adds to

them all irrational numbers and thus develops it into the set of real numbers, *i.e.* into the continuum. We know that all previous additions to the set of numbers were prompted, to greater or lesser degree, by our wish to be able to perform some operations under all possible conditions, which could otherwise not always be achieved with the help of the older system of numbers. Thus the introduction of zero and negative numbers enabled us to deal with all cases of subtraction; the introduction of fractions produced the same result with regard to division (with the exception of division by zero which, by the way, still remains impossible even within our new system of real numbers); the introduction of irrational numbers was prompted by our desire to be able to extract roots. This tendency to deal with operations, which could not always be performed otherwise within the existing set of numbers, was prompted in mathematics not so much by an abstract argument leading to a formal goal (as it is sometimes believed), but by practical necessity; this is best seen by examples like those introduced at the beginning of this chapter; thus we were unable to obtain results in cases where the length of the diagonal of a square of unit side, was required, or, when we were trying to find the area of a circle of unit radius, because the set of numbers at our disposal was insufficient for this purpose.

A strict scrutiny of our principle of origin thus shows that the introduction of the wider set of numbers was prompted by our wish to be able to perform certain operations under all circumstances whereas this could not always be achieved with the help of rational numbers alone. This involves the *creation of a limit of a bounded increasing sequence of numbers*. This is no longer an arithmetical operation. One of the characteristics of an arithmetical operation is that all arithmetical operations are always performed with a *finite* group of numbers; on the other hand, our operation requires the existence of an *infinite* sequence of numbers, and with the help of all these numbers a new number is originated, which is the limit of this sequence. This is an *analytical* operation, *i.e.* one of the first and simplest operations of mathematical analysis.

The widening of the existing set of numbers which was undertaken in order to guarantee performance of any operation with numbers achieves its goal only if the operation is possible within the wider range of numbers. We must therefore convince ourselves that *every bounded increasing sequence has a limit within the range of real numbers*. However, this can readily be proved. In fact let

$$\alpha_1, \alpha_2, \dots, \alpha_n, \dots \quad (6)$$

be one such sequence, *i. e.* $\alpha_{n+1} \geq \alpha_n$ for any n , and there is a number C which is such that $\alpha_n < C$ for any n ; here all α_n are arbitrary real numbers.

If from a certain place onwards, all the numbers in the sequence (6) are equal to one another, then the common value of these numbers will evidently be the limit of the sequence (6); therefore, we can right from the beginning reject this case and assume that the sequence (6) contains an infinite number of different unequal numbers. Let us assume that these different unequal numbers in the sequence (6) increase in the following order

$$\beta_1, \beta_2, \dots, \beta_n, \dots (\beta_{n+1} > \beta_n).$$

Denoting by r_n any rational number between β_n and β_{n+1} *, we have

$$\beta_1 < r_1 < \beta_2 < r_2 < \dots < \beta_n < r_n < \beta_{n+1} < \dots$$

The sequence of the rational number r_n is evidently an increasing bounded sequence with an upper limit, and therefore, according to our principle of origin, it tends to the limit α (which can be rational or irrational); owing to the fact that $r_{n-1} < \beta_n < r_n$, we have, according to theorem 10, § 11, $\beta_n \rightarrow \alpha$ ($n \rightarrow \infty$). But the sequence (6) consists of the same numbers β_n , each of which is, generally speaking, repeated several times; therefore, $\alpha_n \rightarrow \alpha$ ($n \rightarrow \infty$).

We can thus say that our aim to find a region of real numbers (continuum) by extending the region of rational numbers has been achieved; the new operation is one of the basic operations of mathematical analysis which involves transition from an increasing bounded sequence to its limit, and can readily be performed within our extended region of numbers.

This property of continuum is of fundamental importance in the purely logical construction of mathematical analysis as we shall see in later sections.

§ 18. Fundamental Lemmas

The fundamental property of continuum which we have just established above enables us to draw far-reaching conclusions which tend to define more fully and in greater detail the set of real

*The arithmetical theory of real numbers shows that infinitely many rational numbers can be taken between any two real numbers.

numbers, its structure and the laws governing them. We are only interested here in conclusions which can be applied most generally to the structure of mathematical analysis. We shall establish several such theorems in this section; we call them “fundamental lemmas”, because each of them essentially contains one of the most frequently encountered methods of application of continuum to the structure of analysis. Thorough comprehension of these auxiliary propositions which will frequently be referred in future enables us to simplify and abbreviate subsequent treatment of the subject.

Let us understand by a *linear section* a set of all real numbers x which satisfy the inequalities $a \leq x \leq b$, where a and b ($a < b$) are two arbitrary real numbers; we shall assume that this linear section contains both its “ends” a and b ; under such circumstances it is sometimes said to be a “closed” section, in contrast to an “open” section which is defined by the inequality $a < x < b$ (not containing its ends). Let us call the sequence of sections

$$\Delta_1, \Delta_2, \dots, \Delta_n, \dots \quad (1)$$

(contracting if 1) all points of the section Δ_{n+1} belong to the section Δ_n for any n (symbolically $\Delta_{n+1} \subset \Delta_n$), and (2) $\Delta_n \rightarrow 0$ ($n \rightarrow \infty$), where Δ_n denotes the length of the same section.

Lemma 1 (on contraction of a sequence of sections). *If (1) is a contracting sequence of sections, then a single number α exists which belongs to all sections Δ_n .*

Proof Let us denote by a_n and b_n the left and right ends of the section Δ_n respectively; then it is evident that $a_1 \leq a_2 \leq \dots \leq a_n \leq \dots$, and $a_n < b_1$ for any n ; hence the sequence of numbers a_n is an increasing sequence bounded from above, i.e. $\lim_{n \rightarrow \infty} a_n = \alpha$.

Let us now assume that k is an arbitrary natural number; if $n > k$, then the section Δ_n belongs completely to the section Δ_k , so that $a_k \leq a_n \leq b_k$; let us now assume that $n \rightarrow \infty$ and k remains constant; owing to the fact that at the same time $a_n \rightarrow \alpha$, we have from the last inequalities of theorem 9, § 11

$$a_k \leq \alpha \leq b_k,$$

i.e. the number α belongs to the section Δ_k ; also in view of the fact that k is arbitrary, it follows that α belongs to all sections of the given sequence. In order to prove uniqueness of this number let us assume that there is yet another number $\beta > \alpha$ which also belongs to all

sections Δ_k ; in that case the length of each of these sections should not be less than $\beta - \alpha$, which contradicts the condition that $\Delta_k \rightarrow 0$ ($k \rightarrow \infty$). Hence lemma 1 is proved.

Let us now assume that we are given a system of sections (S) (finite or infinite). Let us agree to say that the system (S) covers a certain section Δ if each one of the numbers belonging to Δ lies within at least one of the sections of the system (S).*

Lemma 2 (on finite coverage). *If the system (S) covers the section Δ , then a finite system of section can be chosen from it which would also cover the section Δ .*

Proof. For the sake of brevity we can say that an arbitrary section δ permits finite coverage if it can be covered by a finite group of sections chosen from the system (S). We shall prove converse of lemma 2, i.e. we assume that the section Δ does not permit finite coverage and thus try to show a contradiction. Let us divide the section Δ into halves; if both halves permit finite coverage, then the whole section Δ will evidently also permit finite coverage; but since it does not permit finite coverage, it follows that at least one half of it does not permit finite coverage; let us denote this half by Δ_1 (if neither half permits finite coverage Δ_1 would denote either half). The section Δ_1 which does not permit finite coverage is again divided into halves, and we denote by Δ_2 the half which does not permit finite coverage. We can continue this process *ad infinitum* and obtain a sequence of sections $\Delta, \Delta_1, \Delta_2, \dots, \Delta_n$, none of which permits finite coverage; these sections evidently form a contracting sequence; therefore it follows from lemma 1 that a number α exists which belongs to all those sections. Since α belongs to the section Δ which is covered by the system (S), it follows that α lies within at least one section Δ^* of the system (S). But each of the sections Δ_n contains the number α ; also the length of the section Δ_n tends to zero as n increases, and therefore the section Δ_n will lie completely within the section Δ^* for a sufficiently large n (Fig. 9). This gives us the necessary contradiction: on one hand the section Δ_n does not, by its definition, permit finite

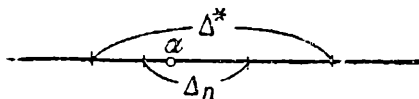


Fig. 9

* We must emphasize the importance of the condition that each number of the section Δ must lie *within* one of the sections of the system (S) and should not simply belong to it; if it were the case, then lemma 2 would not be true.

coverage while on the other it is covered by one of the sections Δ^* of the system (S) . Lemma 2 is thus proved.

We shall now introduce a very important concept of *bounds* of the given set of numbers. The set M of real numbers is said to have an *upper bound* if a number C exists such that all numbers of the set M are smaller than C ; thus the set of all negative numbers has an upper bound (where C can be zero or any positive number); on the other hand the set of all positive numbers has no upper bound. Similarly the set M is said to have a *lower bound* if there is a number C such that all numbers of the set M are greater than C . A set which has an upper and a lower bound is simply said to be *bounded*.

We now say that the number β is the *upper bound* of the set M if this set does not contain numbers which are greater than β , there exists a member of the set which is greater than $\beta - \epsilon$ for an arbitrarily small $\epsilon > 0$. Similarly we say that the number α is the *lower bound* of the set M if this set contains no numbers smaller than α , but there exists a member of the set which is smaller than $\alpha + \epsilon$ for small $\epsilon > 0$. It is thus obvious that the upper bound is the smallest number which does not exceed any number of the set M ; a similar case applies to the lower bound.

Example. The set of positive rational numbers whose squares are smaller than 2 has 0 as its lower bound and $\sqrt{2}$ as its upper bound.

In general, both upper and lower bounds of the set M may or may not belong to this set. The upper and lower bounds of a section evidently coincide with its ends and always belong to it; on the other hand in the above example the set under consideration does not contain its lower bound (since it is not positive), nor its upper bound (since it is not rational).

A set which has no upper limit cannot have an upper bound, for there is no number β in comparison to which all numbers of the given set are smaller. For the purpose of analysis it is important to note that a set with an upper limit always has an upper bound (and only one); similarly, a set with a lower limit always has a single lower bound. The theorem on existence of bounds for bounded sets (which is by no means self-evident) is one of the most important properties of continuum. It can readily be shown that, for example, the set in our last example has an upper limit, but that within the region of rational numbers it has no upper bound.

This theorem is proved in the same way for the upper and lower bounds so that we only need to prove one of these cases.

Lemma 3 (on existence of bounds for bounded sets). *The set M which has an upper limit has a single upper bound.*

Proof. Let us say that a section is *normal* if it contains at least one point of the set M , and to the right of this point there is no point of this set. It can readily be shown that from the two halves of a normal section at least one half will always be normal; in fact, if the right half contains at least one point of the set M , then this right half will evidently be a normal section; if, on the other hand, the right half contains no points of the set M , then the left half will be the normal half.

Let us assume that a is an arbitrary point of the set M , and b is an arbitrary number which exceeds all numbers of the set M . The section $(a, b) = \Delta_1$ is evidently normal; let Δ_2 denote its normal half, Δ_3 the normal half of the section Δ_2 , and generally, Δ_{n+1} the normal half of the section Δ_n ($n = 1, 2, \dots$). The sections $\Delta_1, \Delta_2, \dots, \Delta_n, \dots$ form a contracting sequence, and therefore, according to lemma 1, they have a single point β in common. We now maintain that β is the upper bound of the set M . To begin with, we must prove that there are no points of the set M to the right of β ; let us assume that $\alpha > \beta$ belongs to the set M ; each section Δ_n contains the point β ; but if this is so, it must also contain α , for if it were to end more to the left, then the point α of the set M would lie to its right, and it would no longer be normal. Therefore, each of the sections Δ_n contains both points β and α , and therefore its length is not less than $\alpha - \beta$; however, this is impossible, since $\Delta_n \rightarrow 0$ ($n \rightarrow \infty$). Hence there are no points of the set M to the right of the point β .

Let us now assume that ε is an arbitrary positive number; when n is sufficiently large, $\Delta_n < \varepsilon$; and since Δ_n contains β , all points of the section Δ_n lie to the right of $\beta - \varepsilon$; but since the section Δ_n is normal, it does not contain a single point of the set M to the right of $\beta - \varepsilon$. Therefore, no matter how small $\varepsilon > 0$ there is a point belonging to the set M which lies to the right of $\beta - \varepsilon$; this means that β has also the second property of the upper bound and therefore it is, in fact, the upper bound of the set M ; existence of the upper bound is thus proved. The fact that it is impossible for a given set M to have two different upper bounds is almost self-evident; if there were two such bounds β_1 and β_2 ($\beta_1 < \beta_2$), then it would

follow from the first property of the bound β_1 that no number of the set M can lie between β_1 and β_2 , whereas, according to the second property of the bound β_2 , there must exist such numbers which leads to the required contradiction. Lemma 3 is thus proved.

§ 19. Final Points in Connection with the Theory of Limits

In chapter II we constructed the basic theory of limits. However, some of the more important propositions of this theory could only be established on a more accurate basis which we now have at our disposal after having studied continuum and its fundamental properties. In this section we shall, therefore in a way, supplement our present knowledge of limits.

1. To begin with, let us consider changes in the increasing bounded quantities within a wider scope. If a_n belongs to an increasing sequence of real numbers bounded from above, then a $\lim_{n \rightarrow \infty} a_n$

must exist; (this follows from the last theorem in § 17). But we know that a sequence of numbers is the only way to describe a mathematical process. If we are given an arbitrary process described by any method, we shall naturally say that the quantity x which participates in this process is an increasing quantity if for any two given moments of the process its value at the later moment is not less than its value at the earlier moment. We say that the quantity x has an *upper limit* in the given process if there is a number C , such that from a certain moment of our process onwards we always have $x < C$. It is evident that the increasing sequence with an upper limit, which we have considered at the end of § 17, is a particular case in the general system of increasing quantities with upper limits. We shall see that the theorem at the end of § 17, which was proved for this particular case, remains valid for our general system.

Theorem 1. *Every increasing quantity which is bounded from above has a limit.*

Proof. Owing to the fact that the quantity x is increasing, and bounded from above, there must be a number C such that always $x < C$; therefore, the set M of the values taken by x is bounded from above, and, in accordance with lemma 3 § 18, it has an upper bound β . Let ε be a positive number which can be as small as we please. In accordance with the second property of an upper bound, there must be a number in the set M (i.e. x will sooner or later take this value) which is greater than $\beta - \varepsilon$; since x is an increasing quantity,

all its subsequent values will be greater [than $\beta - \varepsilon$. But it follows from the first property of an upper bound that no number of the set M exceeds β . Hence from a certain moment onward we always have :

$$\beta - \varepsilon < x \leq \beta,$$

and therefore

$$|x - \beta| < \varepsilon;$$

but since the number ε is as small as we please, therefore in this process $x \rightarrow \beta$. This proves theorem 1.

It is evident that this theorem remains valid for a decreasing quantity which is bounded from below.

We have so far said that x increases in the given process if its value x_2 at a later moment of the process is never smaller than its value x_1 at an earlier moment: $x_2 \geq x_1$. Thus an increasing quantity either increases in the course of the process or maintains its former value but it never decreases; therefore, we can naturally say that this quantity is *non-decreasing* and reserve the term "increasing" for quantities for which we always have $x_2 > x_1$ with no possibility of equality. We shall use this terminology in future. Thus, for example, as x increases, the quantity $4x^3$ also increases, but $|x|$ (c.f. § 4, example 1) is a merely non-decreasing quantity. It is obvious that every increasing quantity is at the same time also a non-decreasing quantity, but the converse is not true. Similarly we say that x is a *decreasing* quantity if we always have $x_2 < x_1$ and that it is a *non-increasing* quantity if we always have $x_2 \leq x_1$. All non-decreasing and all non-increasing quantities are called *monotonic* (they always change in the same direction). Hence, in general, theorem 1 can be stated as follows: *a monotonic quantity which is bounded in the direction of its change always has a limit.*

2. Let us now consider a new problem. We have just shown that for a monotonically changing quantity, boundedness in the appropriate direction serves as a sufficient condition for existence of a limit. In general, when a quantity does not change monotonically, it is often important to find if this quantity has a limit in the given process. A necessary and sufficient condition also exists for the general case, as we shall later see and is of great theoretical importance. We formulate and prove this condition for the general case as follows :

Theorem 2 (criterion for existence of a limit). *In order that x should in the given process tend to a limit, it is necessary and sufficient that, no matter how small the positive number ε , from a certain moment of the process onwards, two arbitrary values of x should differ from one another by not less than ε .*

Proof. We shall break up the condition of *sufficiency* into three stages.

1. According to the conditions of the theorem, there will be a moment in our process after which two values of x will differ from one another by less than unity. If at the moment in question $x = x_0$, then for all subsequent moments

$$x_0 - 1 < x < x_0 + 1.$$

Hence the set M of values of x is, from that moment onwards, fully contained in the section Δ with ends $x_0 - 1$ and $x_0 + 1$.

2. Let us call any section δ *normal* if, from an arbitrary moment of the process onwards, x takes values which still belong to the section δ (this can be stated more briefly by saying that the normal section contains values of x which can be "as late as we please"). It is evident that (1) the section Δ is normal, and (2) if the given section is normal, then at least one half of it is also a normal section. This latter circumstance enables us to use the conventional method for constructing a contracting sequence of sections without the use of Δ

$$\Delta, \Delta_1, \Delta_2, \dots, \Delta_n, \dots,$$

in which each section represents the normal half of the preceding section. The common point of all these sections (which exists, as shown by lemma 1, § 18) is denoted by a .

3. Let us prove, finally, that $\lim x = a$. Let ε be an arbitrary positive number. Let n be so large that $\Delta_n < \frac{1}{2}\varepsilon$. Let us fix a moment of our process so that from that moment onwards two arbitrary values of x differ from one another by not less than $\frac{1}{2}\varepsilon$. Since the section Δ_n is normal, it contains a value x_1 which x takes after the moment in question. Hence for any value x_2 taken after that moment, we have, as a result of the choice of that moment,

$$|x_2 - x_1| < \frac{1}{2}\varepsilon.$$

But on the other hand, since both a and x_1 belong to the section Δ_n , whose length is not less than $\frac{1}{2}\varepsilon$ we have

$$|x_1 - a| < \frac{1}{2}\varepsilon.$$

We find from the last two inequalities:

$$|x_2 - a| < \varepsilon;$$

where ε is an arbitrary positive number and x_2 an arbitrary sufficiently late value of x (which it takes after the chosen moment). But this means that $\lim x = a$.

Necessity of our condition can be proved very simply: if $\lim x = a$, then for any two sufficiently late values x_1 and x_2 of x we have

$$|x_1 - a| < \frac{1}{2}\varepsilon, \quad |x_2 - a| < \frac{1}{2}\varepsilon,$$

hence

$$|x_1 - x_2| < \varepsilon,$$

which was to be proved.

The proved condition is very useful in theory, but for proving existence of a limit in individual examples it is rather rarely used; this is due to the fact that in majority of examples, it is rather difficult to determine whether the requirements of this condition are satisfied.*

We have proved the criterion for existence of a limit in very general terms for processes of all types. Evidently when we are trying to apply this criterion to the mathematical structure of a given process, as we did in § 14, the general criterion gives us a definite condition in relation to the process of the given type.

Let us now state other more important special conditions of this type.

1. *In order that the sequence of real numbers $a_1, a_2, \dots, a_n, \dots$ should have a limit, it is necessary and sufficient that the following condition is satisfied: no matter what the positive number ε be, there is a natural number n_0 such that $|a_n - a_m| < \varepsilon$ for $n > n_0, m > n_0$. In other words,*

* This condition is often known as *Cauchy's criterion*; in general it is usual to use the term *criterion* in connection with conditions which are simultaneously necessary and sufficient.

any two "sufficiently far removed" terms of the sequence should differ from one another as little as possible.

2. In order that the function $y = f(x)$ should have a limit for $x \rightarrow a$, it is necessary and sufficient that the following condition is satisfied: no matter what the positive number ε be, there exists another positive number δ such that we always have $|f(x_1) - f(x_2)| < \varepsilon$ for $|x_1 - a| < \delta$ $|x_2 - a| < \delta$ ($x_1 \neq a$, $x_2 \neq a$). In other words, the values of the function $f(x)$ at two different points sufficiently close to a should differ from one another as little as possible.

3. In order that the function $y = f(x)$ should have a limit as x increases indefinitely ($x \rightarrow +\infty$), it is necessary and sufficient that the following condition is satisfied: no matter what the positive number ε be, there exists another positive number A such that we always have $|f(x_1) - f(x_2)| < \varepsilon$ for $x_1 > A$, $x_2 > A$. In other words, the values of the function $f(x)$ for two sufficiently large values of x should differ from one another as little as possible.

Finally, at the end of the section on limits, we find it necessary to say that in order to acquire practice in the evaluation of limits, it is necessary to solve many examples. Many instructive examples of this type can be found in the Problem Book by B.P. Demidovich, in which problem Nos. 38, 40, 41, 42, 48, 50, 50-58, 60, 68, 76, 109-112, 357-365, 376-380 (section I) are particularly useful. *

* At the end of this book the numbers of these exercises are given as they appear in the second edition of the "Problem Book" by B.P. Demidovich.

CHAPTER V

CONTINUOUS FUNCTIONS

§ 20. Definition of Continuity

After the preliminaries we can now study the main problem of mathematical analysis, *viz.* the problem of functional dependence. But even now we must approach our subject systematically and isolate theoretically and practically some classes of functions which are of fundamental importance. Keeping in mind the history of development of our science it is advisable to consider at the beginning the class of *continuous* functions. The concept of continuity, *i.e.* the continuous change of a function, can readily be visualized and we have already used this term on several occasions without having defined its meaning. We must now clearly define the concept of continuity and study the properties of continuous functions in detail, not only because we shall often encounter these functions in future but also because the study of other, more complicated classes of functional relationships can frequently be reduced to the study of continuous functions.

Let $y = f(x)$ be a function which is defined along some section of the number line and let a be an arbitrary point on that line. The function $f(x)$ has a definite value $f(a)$ at the point a . Let us now go from the point a to another adjacent point $a + h$, where h is a positive or negative number with very small absolute value. In connection with this type of transition it is customary to say that the quantity x whose value is a has received an *increment* h and thus taken a new value $a + h$; we have already said that the increment h can be either positive or negative. A new value $f(a + h)$ of the function $f(x)$ corresponds to the new value $a + h$ of x ; the difference $f(a + h) - f(a)$ which corresponds to the difference between the new and old values of y is naturally said to be the *increment* of y which

it has received in the transition of x from the old value a to the new value $a + h$; it is obvious that this increment can be either positive or negative (sometimes it may be zero). In analysis it is customary to denote the increment received by a quantity u by the symbol Δu . We can therefore say that if we have $x = a$, then the increment $\Delta y = f(a + h) - f(a)$ of y corresponds to the increment $\Delta x = h$ of x . Its geometrical meaning is represented in Fig. 10.

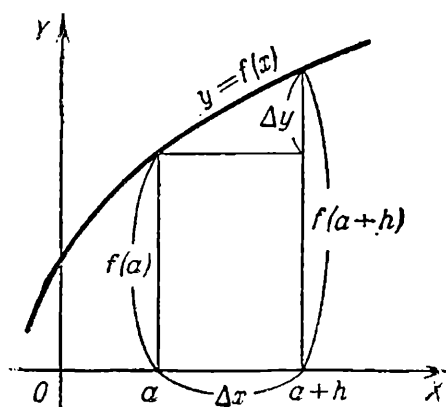


Fig. 10

If a remains unchanged while we change the increment h of x , then evidently the increment $\Delta y = f(a + h) - f(a)$ of y will also change; a definite value of Δy corresponds to each value of Δx . Let us assume that in a particular case the value of h tends to zero, i.e. we assume that the new value of $a + h$ of x tends to the old value a ; if under these circumstances the increment Δy of the function $y = f(x)$ also tends to zero, it would mean that for a sufficiently small change in the value of x the quantity y will also change by as little as we please. This is the meaning of physical representation of the concept

of continuity. Hence the essence of the concept of continuity is the fact that *an infinitely small increment of the function corresponds to an infinitely small increment of the independent variable*. Since the relation

$$\Delta y = f(a + h) - f(a) \rightarrow 0 \quad (\Delta x = h \rightarrow 0)$$

is equivalent to the relation

$$f(a + h) \rightarrow f(a) \quad (h \rightarrow 0),$$

the definition of continuity can be formulated as follows :

The function $f(x)$ is said to be a continuous function at $x = a$ (or "at the point a ") if

$$f(a + h) \rightarrow f(a) \quad (h \rightarrow 0).$$

Hence it is necessary and sufficient for the function $f(x)$ to be continuous at the point a that the value of the function $f(x)$ should tend to a definite limit when $x \rightarrow a$, and that this limit should be equal to the value $f(a)$ of this function at the point a . At the same time it is also important that the relation $f(a + h) \rightarrow f(a)$

should hold *irrespective of the path by which h approaches zero*: by positive values, by negative values or without a change of sign taking place; in other words, we should have $f(x) \rightarrow f(a)$ irrespective of whether x approaches the point a from right or left or whether it passes repeatedly from right to left and *vice versa* (the *two-sided limit* of a function, *cf.* § 15).

The precise definition of the concept of limit transitions which we have studied in detail in § 14 enables us to define the concept of continuity in another way which is often very convenient: *the function $f(x)$ is said to be continuous at the point a if no matter how small $\varepsilon > 0$ there is a $\delta > 0$ such that we have $|f(a+h) - f(a)| < \varepsilon$ for every h whose absolute value is smaller than δ* . In other words, a function is continuous at a given point if a change of the function which can be as small as we please corresponds to a sufficiently small change of the argument.

The majority of cases in which continuity of a function is violated at some point is due to the fact (fig. 11) that $f(x)$ tends to a definite limit when x approaches a from right ($h > 0$) and tends to another definite limit when x approaches a from left ($h < 0$) but these two limits do not coincide. In this case there is no single limit $\lim_{x \rightarrow a} f(x)$ and the function $f(x)$ is discontinuous at the point a as can be readily seen from fig. 11. The fact that x tends to a from right (*i.e.* by assuming values greater than a only) is usually denoted symbolically as follows: $x \rightarrow a + 0$; if in this process $f(x)$ tends to a definite limit, then this limit is denoted by $f(a + 0)$ so that

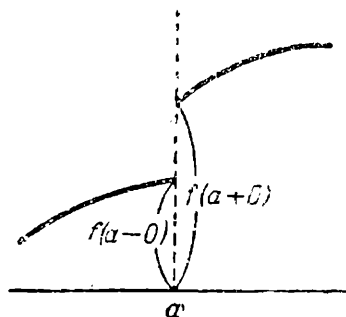


Fig. 11

$$f(a + 0) = \lim_{x \rightarrow a + 0} f(x).$$

The meaning of the symbols $x \rightarrow a - 0$ is similar and

$$f(a - 0) = \lim_{x \rightarrow a - 0} f(x).$$

In the case considered in fig. 11 both limits $f(a + 0)$ and $f(a - 0)$ exist but differ from one another. We know that it is necessary for the function $f(x)$ to be continuous at the point a that not only the

limits should coincide but also each limit should coincide with the value $f(a)$ of the function $f(x)$ at the point a (it is evident from their definition that the numbers $f(a+0)$ and $f(a-0)$ must not necessarily coincide with $f(a)$ and are quite independent of it). Thus apart from the condition $f(a+0) \neq f(a-0)$ which causes violation of continuity in the example we are considering, this phenomenon may also be due to the influence of other causes, *viz.*:

(1) $f(a+0)$ or $f(a-0)$ may not exist at all. Typical examples of this kind are the following:

$$(a) \quad f(x) = \begin{cases} \frac{1}{x} & (x \neq 0), \\ 0 & (x = 0); \end{cases}$$

$f(x)$ increases indefinitely for $x \rightarrow +0$; the absolute value of $f(x)$, which is negative, increases indefinitely for $x \rightarrow -0$; therefore $f(+0)$ and $f(-0)$ do not exist; the function $f(x)$ is thus *unbounded* in the neighbourhood of the point 0.

$$(b) \quad f(x) = \begin{cases} \sin \frac{1}{x} & (x \neq 0), \\ 0 & (x = 0); \end{cases}$$

$f(x)$ remains bounded this time for $x \rightarrow +0$ ($|\sin 1/x| \leq 1$), but cannot tend to a limit, for it repeatedly takes the values $+1$ and -1 (or any value between $+1$ and -1); $f(x)$ evidently behaves similarly for $x \rightarrow -0$; therefore $f(+0)$ and $f(-0)$ do not also exist in this case although the function $f(x)$ remains bounded in the neighbourhood of the point 0.

(2) It may happen that $f(a+0)$ and $f(a-0)$ exist and are equal to one another but differ from $f(a)$; for example

$$f(x) = \begin{cases} x^2 & (x \neq 0), \\ 1 & (x = 0); \end{cases}$$

here $f(a+0) = f(a-0)$ for $a = 0$ while $f(a) = 1$. In all these cases the function $f(x)$ is *discontinuous* (non-continuous) at $x = a$ (at the point a).

It is very important to remember that the definition of continuity implies a *local* (*i.e.* at a given point) property of a function; generally speaking, a function may possess this property at some points while it does not possess it at other points. The values of the

variable x for which the function $f(x)$ is continuous are known as *points of continuity* of this function and points at which the function is discontinuous as *points of discontinuity*. In the examples of discontinuous functions, which we have considered above, the function is continuous everywhere except at a single point; the set of points of discontinuity of such functions evidently consists of this single point. It is quite easy to think of functions which have two, three, or more points of discontinuity and also of functions which have an infinite number of points of discontinuity. But, on the other hand, there are functions which do not have points of continuity at all and points of discontinuity fill the entire number line. An example of this kind is given by the function $D(x)$ which we have considered in § 4; this function is equal to zero or unity in relation to whether x is an irrational or rational number. Owing to the fact that every section of the number line contains an infinite number of both rational and irrational numbers, no matter what the number a be, there will be both rational and irrational numbers in its immediate neighbourhood: hence the function $D(x)$ will take the value 0 and unity at points which can be as close as we please to the point a ; it thus follows that $D(x)$ cannot tend to a limit as $x \rightarrow a$ and it is therefore discontinuous at $x = a$; and owing to the fact that a is arbitrary, the function $D(x)$ is discontinuous everywhere; moreover, $D(x + 0)$ and $D(x - 0)$ do not exist for the same value of x .

It is sometimes useful to distinguish the *one-sided* continuity of a function. The function $f(x)$ is *continuous to right* of the point a if $f(a + 0)$ exists and $f(a + 0) = f(a)$; it is *continuous to left* if $f(a - 0)$ exists and $f(a - 0) = f(a)$; in order that the function should be continuous at the point a it is evidently necessary and sufficient that it should be continuous to right as well as left of that point.

We shall say that the function $f(x)$ is *continuous along the line* $\{a, b\}$ if it is continuous at every point of this line (*i.e.* it has no points of discontinuity on this line). In this case continuity to right of the end a and continuity to left of the end b of this line are only necessary; this fact is obvious because the function $f(x)$ is frequently defined only for points on the line $\{a, b\}$ so that the question of its continuity to left of the point a (or its continuity to right of the point b) does not arise. The definition of a continuous function along a line does not alter our previous statement that continuity is a local property, for continuity along a line is defined in terms of continuity at every point and this is the primary definition in the theory of continuous functions, which implies a well-defined local character.

§ 21. Operations with continuous functions

In chapter 2 we have studied the results of arithmetical operations with infinitely small quantities, infinitely large quantities and quantities which tend to limits; we must now establish the fact that continuity of a function is, as a rule, conserved in the course of elementary arithmetical operations. The importance of this problem is self-evident, for its general solution will make it unnecessary to test continuity of every function obtained as a result of similar operations with continuous functions.

Let us assume that we are given the algebraical sum

$$f(x) = f_1(x) \pm f_2(x) \pm \dots \pm f_n(x)$$

of functions, each of which is continuous at $x = a$. According to the definition of continuity this means that $f_1(x) \rightarrow f_1(a)$, $f_2(x) \rightarrow f_2(a)$, ..., $f_n(x) \rightarrow f_n(a)$ for $x \rightarrow a$; but in this case we know from theorem 1 § 11 that

$$f(x) = f_1(x) \pm f_2(x) \pm \dots \pm f_n(x) \rightarrow f_1(a) \pm f_2(a) \pm \dots \pm f_n(a) = f(a)$$

for $x \rightarrow a$ and this means that the function $f(x)$ is continuous at the point a .

By using a similar simple argument (with reference to theorem 2 § 11) it can be readily proved that the product of an arbitrary constant number of functions which are continuous at the point a will also be continuous at that point; in particular, if the function $f(x)$ is continuous at $x = a$, then the function $\{f(x)\}^n$ possesses the same property, where n is an arbitrary natural number. Division, on the other hand, usually requires some additional explanations. Let $f_1(x)$ and $f_2(x)$ be two functions which are continuous at $x = a$ and let $f_2(a) \neq 0$.

In accordance with our assumption $f_1(x) \rightarrow f_1(a)$ and $f_2(x) \rightarrow f_2(a)$ for $x \rightarrow a$; therefore it follows from theorem 7 § 11 that

$$\lim_{x \rightarrow a} \frac{f_1(x)}{f_2(x)} = \frac{\lim_{x \rightarrow a} f_1(x)}{\lim_{x \rightarrow a} f_2(x)} = \frac{f_1(a)}{f_2(a)},$$

which shows that the function $f_1(x) / f_2(x)$ is continuous at $x = a$; Hence this rule applies provided that $f_2(a) \neq 0$. But if $f_2(a) = 0$, then the expression $f_1(x) / f_2(x)$ is devoid of meaning for $x = a$ and therefore continuity of the quotient has no meaning at all in this case.

All the rules which we have so far established remain valid if we consider a function which is continuous not at a single point but along the entire section of a line; this follows directly from the definition, given in § 20, of continuity of a function along a section. In the case of a quotient the result should evidently be stated as follows: if the functions $f_1(x)$ and $f_2(x)$ are continuous along a section of a line and if $f_2(x)$ does not vanish at any point on that line, then the function $f_1(x)/f_2(x)$ is continuous along that line.

§ 22. Continuity of a composite function

Let y be a function of x , $y = f(x)$, defined along some section of a line (a, b) . Let us denote by M the set of numbers which the function $f(x)$ assumes by running through all the numbers on the line (a, b) . Let some third quantity z be another function of y , $z = \psi(y)$, which is defined for all values of y belonging to the set M . When x takes a definite numerical value on the line (a, b) , then $y = f(x)$ also takes a definite value which belongs to the set M ; but in this case $z = \psi(y)$ also takes a definite numerical value. Hence in the long run a definite value of z corresponds to each value of x along the section (a, b) ; in other words, z is a function of x defined along the section (a, b) . It is convenient to denote this dependence as follows:

$$z = \varphi[f(x)],$$

or by two equations:

$$z = \varphi(y), \quad y = f(x).$$

z is not directly defined in terms of the independent variable x but by means of an "intermediate" function y ; z is defined as a function of y and y as a function of x ; therefore z appears to be a function of x . A function which is given in this way is said to be a *composite function* (or a "function of a function").

Example. Let $y = \cos x$, $z = \log y$; the function $\log y$ is defined only for positive values and therefore we shall restrict ourselves to the study of those values of x for which $y = \cos x > 0$; for example, let $-\pi/4 \leq x \leq +\pi/4$; then $y > 0$ and $\log y$ has a definite meaning. We can write

$$z = \log \cos x \quad \left(-\frac{\pi}{4} \leq x \leq \frac{\pi}{4} \right);$$

we know that this function is very important in the logarithmic solution of trigonometrical equations and detailed tables are published for this function. The following are some other simple examples :

$$z = y^2, \quad y = \sin x, \quad z = \sin^2 x,$$

$$z = \frac{1}{1+y}, \quad y = \sqrt{1+x^2}, \quad z = \frac{1}{1+\sqrt{1+x^2}}$$

(both functions are defined for all values of x).

These examples (like the definition of a composite function) show that the term "composite function" describes no new class of functional dependencies but only a specific *method of defining* a function; the simplest functions can be defined in composite form if this is desirable; thus the function $z = x^4$ can be defined with the help of the relations $z = y^2, y = x^2$ and when given in this form it becomes a composite function.

It is obvious that the form in which a function is defined can be even more composite—it may contain not one but two or more intermediate functions.

Example : The function $v = \log(1 + \sqrt{1+x^2})$ can be given in terms of a chain of relations $v = \log u, u = 1 + z, z = \sqrt{y}, y = 1 + x^2$, i.e. it may contain three intermediate functions u, z and y .

Let $z = \varphi(y), y = f(x)$ and let the function $f(x)$ be defined and continuous along the section (a, b) , while the function $\varphi(y)$ is defined and continuous along another section which includes all values of the function $f(x)$ for $a \leq x \leq b$. We will prove that in this case the composite function $z = \varphi[f(x)]$ is also continuous along the line (a, b) . Let α be an arbitrary point on the line (a, b) , and let $f(\alpha) = \beta$; since the function $f(x)$ is continuous along the line (a, b) , we have :

$$\lim_{x \rightarrow \alpha} f(x) = f(\alpha) = \beta;$$

and, on the other hand, the function $\varphi(y)$ is assumed to be continuous for $y = \beta$; it therefore follows from $f(x) \rightarrow \beta$ that

$$\varphi[f(x)] \rightarrow \varphi(\beta) = \varphi[f(\alpha)] \quad (x \rightarrow \alpha),$$

which proves continuity of the composite function $\varphi[f(x)]$ at the point α ; and since α is an arbitrary point on the line (a, b) , the function $\varphi[f(x)]$ must be continuous along the whole line. We have thus proved the following theorem.

Theorem 1. *If the function $f(x)$ is continuous along the line (a, b) and the function $\varphi(y)$ continuous along another line which contains all values of the function $f(x)$ along the line (a, b) , then the composite function $\varphi[f(x)]$ is continuous along the line (a, b) .*

In other words, if both functional dependencies which make a composite function are continuous, the composite function will also be continuous. By means of simple induction it is easy to apply this theorem to composite functions defined in terms of three or more links: if in each link the dependence is continuous, then the resulting composite function will be a continuous function. In practice we always come across composite functions, each link of which is composed of some elementary function (c f. § 6). It would indeed be very difficult to prove in each case the continuity of each combination of elementary functions which we may encounter. Theorem 1 enables us once and for all to dispense with this necessity: if we can prove continuity of a small number of the simple elementary functions (and we shall do this in § 24), then it follows from theorem 1 and from the theorems in § 21 that every finite combination of these simple functions will be continuous (*i.e.* every such combination composed of the simple elementary functions by means of arithmetical operations and operations involved in constructing a composite function, which can be repeated an arbitrary finite number of times in any order).

§ 23. Fundamental properties of continuous functions

Continuous functions have a series of properties which make their study and application much simpler than is the case with non-continuous functions. We shall now state and prove several important properties of this kind. But, to begin with, we must establish an auxiliary proposition which we shall find very useful in future.

Lemma. *The function $f(x)$ which is continuous and positive for $x = a$ will also be positive along some line which contains the point a .*

Proof. As a result of continuity of the given function at the point a we have:

$$f(x) \rightarrow f(a) \quad (x \rightarrow a);$$

and therefore it follows from $f(a) > 0$ and theorem 2 § 10 that $f(x) > 0$ provided x is sufficiently close to a ; this proves the statement of the lemma.

Obviously it is possible to prove by the same method that if $f(a) < 0$, we should have $f(x) < 0$ for all points on some line which contains the point a .

In future we shall say that the function $f(x)$ defined along the line (a, b) is bounded along that line if its values along the line (a, b) form a bounded set.

Theorem 1. *The function $f(x)$ continuous along the line (a, b) is bounded along that line.*

Proof. Let α be an arbitrary point on the line (a, b) . The function $f(x)$ is continuous at the point α , and therefore $|f(x) - f(\alpha)| < 1$ provided x is sufficiently close to α ; hence a line δ_α exists with α as its centre such that for an arbitrary point x on the line δ_α *) we have $|f(x) - f(\alpha)| < 1$; therefore

$$|f(x)| < |f(\alpha)| + 1.$$

We can construct a line δ_α for every point α on the line (a, b) . The system S of all the lines so constructed evidently covers the line (a, b) . In accordance with the theorem on finite coverage (lemma 2 § 18) a finite group of lines $\Delta_1, \Delta_2, \dots, \Delta_n$ of the system S exists which also covers the line (a, b) . Each one of the lines Δ_k is one of the constructed lines δ_α ($k = 1, 2, \dots, n$); therefore for any point x on the line Δ_k we have

$$|f(x)| < |f(\alpha_k)| + 1;$$

If we denote by μ the smallest number among n numbers $|f(\alpha_1)|, |f(\alpha_2)|, \dots, |f(\alpha_n)|$ and keep in mind the fact that any point x on the line (a, b) belongs to at least one of the lines Δ_k , then we have for any point x on the line (a, b)

$$|f(x)| < \mu + 1.$$

This proves that the function $f(x)$ is bounded along the line (a, b) .

Theorem 2. *The function $f(x)$ continuous along the line (a, b) takes its minimum and maximum values on that line.*

Preliminary note. It follows from the above theorem that the function $f(x)$ is bounded along the line (a, b) ; it follows from lemma

*) It is obviously assumed that x lies on the line (a, b) ; if the point α coincides with one of the ends of the line (a, b) , then the inequality $|f(x) - f(\alpha)| < 1$ must only be satisfied for those points x on the line δ_α which lie on the line (a, b) .

§ 18 that the set M of values taken by the function $f(x)$ along the line (a, b) has therefore a lower bound α and an upper bound β . We know, however, that the bounds of a bounded set must necessarily belong to it; hence in the case under consideration α and β may or may not belong to the set M , *i.e.* they must not necessarily be values which the function $f(x)$ takes along the line (a, b) . This may be quite possible for non-continuous functions; let us assume, for example, that

$$f(x) = \begin{cases} x & (x < 1), \\ 0 & (x = 1), \end{cases}$$

if $x < 1$ and is sufficiently close to unity, then the function $f(x)$ will also be as close to unity as we please; the upper bound $\beta = 1$; however, at no point on the line $(0, 1)$ do we have $f(x) = 1$, and so everywhere $f(x) < 1$. Theorem 2 tries to establish the fact that this position is impossible for a continuous function: here the upper and lower bounds of the set M are always the maximum and minimum values, respectively, of the function $f(x)$ along the line (a, b) ; in other words, a point x_1 can always be found on the line (a, b) such that $f(x_1) = \alpha$, and another point x_2 such that $f(x_2) = \beta$.

Proof. We shall give the proof only for the upper bound β , as the argument is exactly the same for the lower bound. Let us assume that $f(x) < \beta$ at an arbitrary point x on the line (a, b) and let us try to arrive at a contradiction. Owing to the fact that the function $\beta - f(x)$ is continuous and does not vanish on the line (a, b) , it therefore follows from the last result of § 21 that the function $1 / \{\beta - f(x)\}$ is a continuous function and, as a result of theorem 1, it is bounded on that line. Thus a number $C > 0$ exists such that

$$\frac{1}{\beta - f(x)} < C \quad (a \leq x \leq b),$$

and consequently

$$f(x) < \beta - \frac{1}{C} \quad (a \leq x \leq b).$$

Since C is a constant positive number, this contradicts the definition of the upper bound, according to which values of $f(x)$ can be found on the line (a, b) , which are greater than $\beta - \varepsilon$ for arbitrarily small $\varepsilon > 0$. This contradiction proves theorem 2.

Theorem 3. *If the function $f(x)$ is continuous on the line (a, b) and if γ denotes an arbitrary number between $f(a)$ and $f(b)$, then a point c can be found between a and b such that $f(c) = \gamma$.*

Preliminary note. Theorem 3 expresses that property of continuous functions which forms in our visual representation the essence of the continuous change ; in passing from one value to another a continuously changing quantity must inevitably run through all intermediate values without omitting any of them.

Proof. At first let us consider the particular case when $f(a)$ and $f(b)$ have opposite signs and $\gamma = 0$ (i.e. we will show that when a continuous function changes from positive to negative or vice versa it must pass through zero). Let us assume that $f(x)$ does not vanish anywhere on the line (a, b) and let us thus try to arrive at a contradiction. Let us agree to call an arbitrary section of a line *normal* if the function $f(x)$ has opposite signs on its ends : it is evident that out of two halves of a normal section one, and only one, will always be normal (let us recall that according to our assumption $f(x)$ does not vanish anywhere). It is given that the line $\Delta_1 = (a, b)$ is normal ; let Δ_2 be the normal half of Δ_1 , Δ_3 the normal half of Δ_2 , and so on. The lines $\Delta_1, \Delta_2, \dots, \Delta_n, \dots$ form a contracting sequence and therefore have a point in common which we can denote by c . According to our assumption $f(c) \neq 0$; therefore either $f(c) > 0$ or $f(c) < 0$. Let us assume that $f(c) > 0$. Hence as a result of the lemma established at the beginning of this paragraph we should have $f(x) > 0$ for all values of x sufficiently close to c : however, this contradicts the fact that the point belongs to the normal section Δ_n which can be as short as we please and hence in the immediate neighbourhood of the point c another point can be found (one of the ends of the section Δ_n) where $f(x) < 0$. This proves the particular case of theorem 3.

Let us now consider the general case and assume that $f(x) - \gamma = \varphi(x)$. According to the conditions of the theorem γ lies between $f(a)$ and $f(b)$; therefore $\varphi(a)$ and $\varphi(b)$ have opposite signs. And since both $f(x)$ and $\varphi(x)$ are continuous functions, it follows from the particular case proved above that a point c can be found between a and b such that $\varphi(c) = 0$ or, which comes to the same thing, $f(c) = \gamma$. This proves theorem 3 completely.

Let us now assume that the function $y = f(x)$ is a continuous increasing function on the line (a, b) ; this means that we always have $f(x_1) < f(x_2)$ for $a \leq x_1 < x_2 \leq b$. Let $f(a) = \alpha, f(b) = \beta$

($\alpha < \beta$) and let γ be an arbitrary number lying between α and β . According to theorem 3 a number c exists ($a < c < b$) such that $f(c) = \gamma$; since the function $f(x)$ is an increasing function, the number c must evidently be unique; thus a definite number c on the line (a, b) corresponds to every number γ on the line (α, β) ; in other words, a single definite value of x on the line (a, b) , for which $y = f(x)$, corresponds to every value of y on the line (α, β) ; hence x is a function of y and is defined on the line (α, β) :

$$x = \varphi(y) \quad (\alpha \leq y \leq \beta);$$

and it is evident that $\varphi(\alpha) = a$, $\varphi(\beta) = b$. The function $x = \varphi(y)$ is said to be inverse of the function $y = f(x)$; these two functions essentially express the same functional dependence between x and y and differ from one another only by the fact that one quantity is taken as the independent variable while the other is the function.

Example 1. $y = x^3$ ($-\infty < x < +\infty$); the inverse function is $x = \sqrt[3]{y}$ ($-\infty < y < +\infty$).

Example 2. $y = \sin x$ ($0 \leq x \leq \pi/2$); the inverse function is $x = \arcsin y$ ($0 \leq y \leq 1$).

Let us now prove that a function which is inverse of an increasing continuous function is also continuous along the corresponding line.

Theorem 4. *Let the function $y = f(x)$ be an increasing continuous function on the line (a, b) and let $f(a) = \alpha$, $f(b) = \beta$; then the inverse function $x = \varphi(y)$ will also be continuous on the line (α, β) .*

Proof. At first let us assume that γ is an arbitrary interior point on the line (α, β) ; let $\varphi(\gamma) = c$ so that $a < c < b$ and $f(c) = \gamma$. Let $\varepsilon > 0$ be so small that the numbers $c - \varepsilon$ and $c + \varepsilon$ lie on the line (a, b) . Assume that

$$f(c - \varepsilon) = \gamma_1, \quad f(c + \varepsilon) = \gamma_2,$$

so that $\gamma_1 < \gamma < \gamma_2$ and denote by δ the smaller of the differences $\gamma - \gamma_1$, $\gamma_2 - \gamma$.

Let us now assume that $|y - \gamma| < \delta$; in this case evidently $\gamma_1 < y < \gamma_2$ and therefore

$$\varphi(\gamma_1) < \varphi(y) < \varphi(\gamma_2);$$

but $\varphi(\gamma_1) = c - \varepsilon$, $\varphi(\gamma_2) = c + \varepsilon$ and consequently

$$c - \varepsilon < \varphi(y) < c + \varepsilon,$$

or what comes to the same thing, $|\varphi(y) - c| = |\varphi(y) - \varphi(\gamma)| < \varepsilon$. We have thus proved that $|y - \gamma| < \delta$ implies that $|\varphi(y) - \varphi(\gamma)| < \varepsilon$; since $\varepsilon > 0$ can be as small as we please, the function $\varphi(y)$ is continuous at the point γ , which was to be proved.

In case $\gamma = \alpha$ or $\gamma = \beta$, the same arguments can be used to establish continuity of the function $\varphi(y)$ to right (at the point α) or to left (at the point β).

It is obvious that the theorem is also valid for a decreasing continuous function $f(x)$.

In § 20 we repeatedly emphasized the local character of our concept of continuity; we said that continuity applies at every individual point so that, generally speaking, a function can be continuous at some points and discontinuous at other points. We have then defined continuity along a line as continuity at every point on the line. However, it is possible to define continuity along a line directly without using the concept of continuity at a point. In doing this we shall also use the basic idea that the essence of continuity is due to the smallness of changes of the function when the independent variable undergoes small changes.

We shall say that the function $f(x)$ is *uniformly continuous* along the line (a, b) if the absolute value of the difference of its values at two sufficiently close points on this line is as small as we please. More accurately: *we say that the function $f(x)$ is uniformly continuous along the line (a, b) if the following condition is satisfied: no matter what $\varepsilon > 0$ be, there exists a $\delta > 0$ such that for two points x_1 and x_2 on the line (a, b) which have a distance $|x_1 - x_2| < \delta$ we have $|f(x_1) - f(x_2)| < \varepsilon$.* We say in this definition that continuity is *uniform* because the difference $|f(x_1) - f(x_2)|$ should be small irrespective of the positions of the points x_1 and x_2 on the line (a, b) , provided these points are situated close to one another.

The concept of uniform continuity is very important in mathematical analysis. We must therefore establish right from the beginning the relationship between this concept and the concept of continuity along a line which we have defined earlier. It is almost self-evident that uniform continuity of a function along a line implies its continuity at every point on that line and hence also its continuity along this line in accordance with our earlier definition. In fact, if the function $f(x)$ is uniformly continuous along a line and if α is an arbitrary point on that line, then the difference $|f(x) - f(\alpha)|$ will

be as small as we please provided x is sufficiently close to α , and this implies continuity of the function $f(x)$ at the point α . It is even more important to note that the converse theorem also holds: the fact that a function is continuous at every point of a (closed) line is *sufficient* for it to be uniformly continuous along this line. Thus our new definition of continuity appears (for closed curves) to be totally equivalent to our former definition (of local character).

Theorem 5. *The function $f(x)$ continuous at every point of the line (a, b) is uniformly continuous along that line.*

Proof. Let α be an arbitrary point on the line (a, b) and let ε be an arbitrary positive number. Since the function $f(x)$ is continuous at the point α , therefore for a sufficiently small $\delta_\alpha > 0$ we have for any point $x^*)$ on the line $(\alpha - \delta_\alpha, \alpha + \delta_\alpha)$

$$|f(x) - f(\alpha)| < \frac{\varepsilon}{2};$$

therefore if x_1 and x_2 are two arbitrary points on the line $(\alpha - \delta_\alpha, \alpha + \delta_\alpha)$, then

$$|f(x_1) - f(x_2)| < \varepsilon. \quad (1)$$

The same construction can be repeated for every point α on the line (a, b) and we can say that the section $(\alpha - \frac{1}{2}\delta_\alpha, \alpha + \frac{1}{2}\delta_\alpha)$, which comprises of one half of the section constructed above is the "proper section" of the point α . The set S of all such "proper sections" evidently covers the whole line (a, b) ; according to the lemma on finite coverage (lemma 2 § 18) a finite group $\Delta_1, \Delta_2, \dots, \Delta_n$ can be chosen from these "proper sections", which will also cover the line (a, b) . Let us denote by δ half the length of the shortest of the sections $\Delta_1, \Delta_2, \dots, \Delta_n$.

Let us now assume that x_1 and x_2 are two arbitrary points on the line (a, b) with a distance not greater than δ . The point x_1 , like every other point on the line (a, b) , lies on a section Δ_k ; but Δ_k is one of the sections of the system S and is therefore a "proper section" $(\alpha - \frac{1}{2}\delta_\alpha, \alpha + \frac{1}{2}\delta_\alpha)$ of a point α on the line (a, b) ; therefore $|x_1 - \alpha| \leq \frac{1}{2}\delta_\alpha$; but on the other hand

$$|x_2 - x_1| < \delta \leq \frac{1}{2}\delta_\alpha,$$

*) We are obviously only considering points x which lie on the line (a, b) , for the function $f(x)$ may be undefined outside this line.

where the above inequality follows from the definition of the number δ . Hence the two numbers α and x_2 have a distance not greater than $\frac{1}{2} \delta_\alpha$ and consequently

$$|x_2 - \alpha| < \delta_\alpha.$$

It therefore follows that both points x_1 and x_2 lie on the line $(\alpha - \delta_\alpha, \alpha + \delta_\alpha)$ and the inequality (1) holds; but x_1 and x_2 are two arbitrary points which have a distance not greater than δ ; hence uniform continuity of the function $f(x)$ along the line (a, b) has been established.

Note. We are assuming, as always, that the line (a, b) which we have considered in theorem 5, is a closed line, *i.e.* it contains both its ends. For open lines (which do not contain their ends) theorem 5 is, generally speaking, not valid. Thus, for example, the function $f(x) = 1/x$ which is continuous at every point of the open line $(0, 1)$ is not uniformly continuous along that line; in fact, for any small $\delta > 0$ and assuming that $x_1 = \delta$, $x_2 = 2\delta$, we have $|x_1 - x_2| = \delta$, $|f(x_1) - f(x_2)| = 1/\delta - 1/(2\delta) = 1/(2\delta)$; although $|x_1 - x_2|$ can be as small as we please, $|f(x_1) - f(x_2)|$ is as large as we please. The function $f(x) = \tan x$ behaves similarly along the open line $(-\pi/2, +\pi/2)$. In both cases we are dealing with unbounded functions. The function $\sin 1/x$, considered in § 20, is continuous at every point on the open line $(0, 1)$ and is evidently bounded along that line; however, it is not uniformly continuous, for the numbers x_1 and x_2 exist and can be as small as we please (and therefore also as close as we please), for which $\sin 1/x_1 = 1$, $\sin 1/x_2 = -1$.

§ 24. Continuity of elementary functions

We shall show in this paragraph that all elementary functions are basically continuous (*i.e.* with the exception of some easily distinguishable points).

1. Theorem 5 § 11 maintains that every polynomial $P(x)$ is continuous for every value of x . Similarly the corollary of theorem 7 in the same paragraph states that every rational fraction $P(x)/Q(x)$ is continuous for every value of x provided its denominator does not vanish. Hence *all rational functions are essentially continuous*.

2. Let us consider the exponential function $y = a^x$ and assume that $a > 1$. Since

$$a^{x+h} - a^x = a^x (a^h - 1),$$

therefore in order to prove continuity of the function a^x for every value of x it is sufficient to show that

$$a^h - 1 \rightarrow 0 \quad (h \rightarrow 0).$$

For this purpose we must at first note, that Newton's binomial

$$(1 + \lambda)^n = 1 + n\lambda + \dots,$$

where $\lambda > 0$ and n is an arbitrary natural number, gives *) for $n > 1$:

$$(1 + \lambda)^n > 1 + n\lambda,$$

hence

$$\lambda < \frac{(1 + \lambda)^n - 1}{n}.$$

Assuming that $h > 0$ and $\lambda = a^h - 1$ in this inequality we find:

$$a^h - 1 < \frac{a^{nh} - 1}{n} \quad (h > 0, n > 1). \quad (1)$$

Let us now choose the number n such that

$$n \leq \frac{1}{h} < n + 1;$$

when $nh \leq 1$, $a^{nh} \leq a$ for $a > 1$; hence the function a^x is an increasing function (cf. § 17); it therefore follows from (1) that

$$a^h - 1 < \frac{a - 1}{n};$$

and since evidently $n \rightarrow \infty$ for $h \rightarrow 0$, therefore,

$$a^h - 1 \rightarrow 0 \quad (h \rightarrow 0),$$

which was to be proved; thus *all exponential functions a^x are continuous for every value of x .*

3. We saw in § 17 that the function a^x is always monotonic for $a > 0$ **): it is an increasing function for $a > 1$ and a decreasing function for $a < 1$. We have already proved continuity of this function; a single inverse of the function a^x must exist and it follows from theorem 4 § 23 that this inverse function must also be continuous

*) we saw this already in example 3 § 7.

**) A function is said to be monotonic if it is either non-decreasing or non-increasing along the given line (which can be the whole real axis).

along the half line $x > 0$. This inverse function is the function $\log a^x$. Hence *all logarithmical functions are continuous*.

4. Since an exponential function is monotonic, it follows that every power function x^α is continuous for every constant index α along the half line $x > 0$. In fact, let us assume that $\alpha > 0$ and let n be an arbitrary integer greater than α . We have :

$$(x + h)^\alpha - x^\alpha = x^\alpha \left[\left(1 + \frac{h}{x} \right)^\alpha - 1 \right] \quad (2)$$

Let us assume that $h > 0$. Since an exponential function is monotonic, we have

$$1 < \left(1 + \frac{h}{x} \right)^\alpha < \left(1 + \frac{h}{x} \right)^n$$

(we are, of course, assuming that $x > 0$); $(1 + h/x)^n$ is a polynomial in h and it evidently tends to unity for $h \rightarrow 0$; therefore it follows from theorem 10 § 11, that for $h \rightarrow 0$

$$\left(1 + \frac{h}{x} \right)^\alpha \rightarrow 1.$$

hence from (2)

$$(x + h)^\alpha - x^\alpha \rightarrow 0 \quad (h \rightarrow 0),$$

which was to be proved. Thus *every power function x^α is continuous for $x > 0$ **.

5. It is very easy to prove continuity of the functions $\sin x$ and $\cos x$ along the whole number line. In fact

$$\sin(x + h) - \sin x = 2 \cos \left(x + \frac{h}{2} \right) \sin \frac{h}{2},$$

where the last factor, and therefore the right-hand side as a whole, tends to zero for $h \rightarrow 0$; the proof is similar for $\cos x$. Finally the functions $\tan x$, $\cot x$, $\sec x$ and $\operatorname{cosec} x$ are expressed in terms of ratios of functions such as unity, $\cos x$ and $\sin x$; hence *all simple trigonometrical functions are continuous at all points for which they are defined*.

*) If the function x^α is written in the form $e^{\alpha \log_e x}$, it can readily be seen, from theorem 1 § 22 that its continuity follows directly from continuity of exponential and logarithmic functions.

6. The application of the theory of inverse functions (theorem 4 § 23) inevitably leads to the conclusion that *all inverse trigonometrical functions are continuous within appropriate regions*. Let us consider, for example, the function $\arcsin x$; since the continuous function $\sin x$ is monotonic along the line $(-\pi/2, +\pi/2)$ and increases from -1 to $+1$, its inverse function is also monotonic and continuous along the line $(-1, +1)$ and increases from $-\pi/2$ to $+\pi/2$; and this inverse function is no other than $\arcsin x$.

This concludes the proof of continuity for all simple elementary functions. We know (§ 6) that all other elementary functions can be obtained from these functions as a result of finite number of algebraic operations; since all these operations are performed with continuous functions, therefore, according to the theorems in §§ 21-22, they result in other continuous functions; this establishes the fact that all elementary functions are continuous everywhere except at isolated points whose positions can be determined in each case from the analytical expression of the function in question.

Example. The function

$$y = \tan \frac{x}{x-1}$$

is continuous everywhere except (1) at the point $x = 1$ and (2) at points x for which

$$\frac{x}{x-1} = (2k+1) \frac{\pi}{2},$$

where k is an arbitrary integer, *i.e.* at points

$$x = \frac{(2k+1)\pi}{(2k+1)\pi - 2} \quad (k = \dots, -2, -1, 0, 1, 2, \dots);$$

hence continuity is violated only at points where the analytical expression of the given function becomes void.

The reader will be able to find exercises relating to Chapter 5 in the Problem Book by B.P. Demidovich, Section I, § 7. We recommend problem Nos. 490-501, 515-518, 544, 566, 568; the choice of other exercises is left to the teacher.

CHAPTER VI

DERIVATIVES

§ 25. Uniform and non-uniform variation of functions

When we study variation of functions in practice we are most interested in the problem of *speed*, *i.e.* in the *rate* at which the phenomenon in question changes. The velocity of movement of a railway carriage or an aeroplane is the main index of its activity. The rate of increase of population of a town is one of its main lively characteristics. A road which rises from lower to higher places can be more or less steep in relation to the rate with which its gradient rises.

The fundamental concept of speed is self-evident. However, for the solution of a majority of practical problems this general concept is insufficient. It is necessary to have an accurate quantitative definition for this quantity which we call the rate of change of the given phenomenon. However, when trying to form such a definition, we find that the methods of elementary mathematics are insufficient and only adequate for a few simpler cases. Generally speaking, this problem can only be solved satisfactorily with the help of some mathematical methods and concepts which we shall now study. In the historical development the general necessity for the accurate definition of the rate of change of quantities and establishment of a unique method for evaluation of this rate led to the development of the science which we call mathematical analysis. An extensive branch of mathematical analysis is devoted to the solution of this problem and its consequences. This branch is usually known as *differential calculus* and we shall now begin its study.

Let us assume that y is a function of (the independent variable) x :

$$y = f(x).$$

The change (increment) Δx of x leads to a definite increment

$$\Delta y = f(x + \Delta x) - f(x) \quad (1)$$

of y . This increment can be very diverse in relation to the initial value x of the independent variable and the nature of the function $f(x)$ which expresses the functional dependence in question; in other words, it can readily be seen from (1) that the increment Δy , apart from the increment Δx , also depends on x and on the form of the function $f(x)$. We naturally assume that y changes *quickly* if $|\Delta y|$ (for a given Δx) is large, and *slowly* if it is small; if $\Delta y = 0$, then y does not change at all as the independent variable x passes from x to $x + \Delta x$.

Let us now consider the simple case when the change Δy of y is always proportional to the change Δx of x , *i.e.* $\Delta y = \alpha \Delta x$, where α is a constant independent of x and Δx . If $\Delta x = 1$, then $\Delta y = \alpha$, *i.e.* the change of x by a unit (of this quantity) always corresponds to the same change $\Delta y = \alpha$ of y irrespective of the initial value x of the independent variable. If x denotes time, then y changes by the same quantity α *during every unit of time* (for example, during every second) irrespective of the moment x when we began to count. In other words, during the whole course of the process y undergoes the same change α in every *unit of time*. We can thus clearly see that in this case changes in y are not accelerated or retarded in the course of the process but always take place at the same rate, *i.e.* we say that y changes *uniformly*. It is evident that changes of this type occur very frequently and therefore this is one of the most important cases; it is of the greatest significance in all that follows and we must therefore consider it in great detail.

Let us assume that $y = f(x)$ changes uniformly and let $y = f(a) = b$ for $x = a$; in that case we have for every x

$$f(x) - f(a) = \alpha(x - a),$$

and therefore

$$f(x) = \alpha x + f(a) - \alpha a = \alpha x + \beta,$$

where $\beta = f(a) - \alpha a$ is a constant. Thus every uniformly changing quantity $y = f(x)$ represents a linear function (a binomial of the first degree) of x :

$$y = \alpha x + \beta. \quad (2)$$

Conversely, if $y = f(x)$ is connected with the independent variable x by the relation (2), then

$$\Delta y = f(x + \Delta x) - f(x) = [\alpha(x + \Delta x) + \beta] - [\alpha x + \beta] = \alpha \Delta x,$$

i.e. Δy is proportional to Δx and y changes uniformly. Thus all linear functions, and only linear functions, change uniformly; this clearly shows that uniform change of functions is a very restricted particular case.

If x denotes the time which elapsed from a certain initial moment and y denotes the distance of the moving body at the moment x from a certain initial position, then the increment Δy of this distance for $\Delta x = 1$ evidently denotes the path covered by the body in unit time (for example in one second). If y changes uniformly, the body would travel the same distance in every second; if this distance is α , then, as we know, $y = \alpha x + \beta$, where β is a constant. In physics this type of movement is called *uniform* and the path α covered by the body in unit time of uniform motion is said to be *velocity* of this uniform motion.

Similarly in the general case, when x and y are representing arbitrary quantities, the uniform change of the function $y = f(x)$ means that the change in x by one unit causes an increment in y equal to one and the same number α ; we naturally assume that in this case the number α also measures the velocity of change of y (*in relation to x*). Hence the definition of velocity of a uniformly changing function causes no difficulties; if such a function is written in the form $y = \alpha x + \beta$, then α is the (constant) rate of its change. α can be positive, negative or zero. If $\alpha < 0$, then $\Delta y < 0$ for $\Delta x > 0$, *i.e.* as x increases, y decreases. This case is obviously quite possible. Thus in the case of uniform movement which we have considered above y can denote not only the distance of the moving body from some initial position but also its distance from its final position which it approaches. In that case y will evidently decrease in the course of time, *i.e.* we shall have $\Delta y < 0$ for $\Delta x > 0$. When $\alpha = 0$, we have $y = \beta$; thus y remains constant in the course of the process; this means in our mechanical example that the body remains at rest, *i.e.* the velocity of its movement is equal to zero.

It is obvious that if a function changes non-uniformly, *i.e.* if the increment of the function acquired for every unit increment of the independent variable is different at different moments of the

process, we cannot so simply define the rate of change of the function. But we can then forecast that in this case the rate of change of the function for any feasible definition of the function will be different at different moments of the process, *i.e.* it will be a *local* phenomenon, where the word “local” has the same meaning as in the last chapter.

§ 26. Instantaneous velocity of non-uniform movement

Let us assume that we are interested in the problem of velocity with which a given car moves at a given moment. This question is often answered as follows : “velocity of the car is 40 km per hour”. But what does this mean? Can it mean that the car travels 40 km per hour? But we are really interested in the speed of the car at a *given moment*, since in the course of one hour the car will probably change its velocity many times by accelerating and retarding ; and if we know that during one hour it travels 40 km, this does not tell us anything at all about the velocity with which it moves now, at this instant.

It may be argued that problem of instantaneous velocity is unimportant and that it is only important to know the total distance travelled by the car in one hour. But this is not true. Let us assume that the car crosses a bridge and a street sign shows that the speed limit must not exceed 10 km. per hour. A policeman stops the driver and fines him for exceeding the speed limit : it appears that the car travelled with the velocity of 20 km per hour. But how do we know this? Who knows the distance covered by the car during the last hour? It is obvious that in this case it is quite unimportant to know the distance the car covered or is going to cover during a given time interval ; we only need to know its velocity now, at this instant.

If the speed of the car is uniform, *i.e.* if it always travels with the same velocity, the assessment “40 km per hour” fully describes its velocity which is one and the same at every moment of its movement. But the car does not move uniformly ; during an hour its velocity changes many times and when we are told that the car covered a distance of 40 km in one hour, this only gives us an impression of *average* velocity of the car at a particular moment at a particular place on its route. An hour is rather a long time interval during which the velocity of the car may change many times.

This will evidently make us choose a smaller time unit than an hour, for example one second. If we assume that during one second

our car covered 20 m, then is that not a sufficient indication of the velocity of its movement, say, during the beginning of that second? It is obvious that in this case the position is much better: as a rule a car will not change its velocity many times during one second and at different moments of this second it will move with more or less the same velocity; therefore, in all probability we may consider the average velocity of the car in one second to be a good approximate assessment of its "instantaneous" velocity of movement at every instant of that second.

This, then, is the state of affairs with a rough object like a car. However, in physical and technical problems various more accurate cases are met with, when during one second a moving body is able to change its velocity more frequently and within much wider limits than a car is able to do within an hour. Just imagine one of the minutest particles of matter, an atom or an electron, which is subjected to milliards of collisions every second, each of which radically affects its velocity. It is obvious that for such a minute particle a second will be an enormous historical era in its life and that the path covered by the particle during one second will tell us nothing about the speed with which the particle moved at a given moment.

We must therefore study the problem from a general point of view and we are now ready to do this. Let us denote by t the time which elapsed from a certain chosen instant, which will be the same once and for all, and by s the path covered by the moving body from the initial instant until the instant t^*). A definite value of s corresponds to every value of t so that s is a function of t :

$$s = f(t).$$

This equation is usually called the *law of motion* of the given body; we shall assume that we are already familiar with this law.

The problem in which we are interested is as follows: how can we find the velocity of a moving body at a given instant of time t from the knowledge of its law of motion? Before we try to solve this problem, we must make a very important methodical remark which is necessary for correct understanding of the given problem

*) For the sake of simplicity it is useful to assume that the body moves along a straight line; however, all that follows remains valid when much wider assumptions are made.

and which will be equally applicable to a great majority of practical problems which we are going to consider in future.

In a majority of ordinary problems when we are trying to evaluate some quantity (the square root of a quadratic equation, the length of a side adjacent to a right angle in a right-angled triangle, *etc.*) we know well in advance the nature of the required quantity, *i e.* we know its general definition; we only need to evaluate its numerical value or its symbolic expression, depending on the character of the given problem.

In this case the position is completely different: what do we mean by the velocity of a moving body at the given instant t — we do not know this, we did not define anywhere its meaning; it may at first appear that the problem in question is insolvable; how, in fact, should we proceed to evaluate a quantity about which we do not even know what it represents and defines? In order to make our problem feasible and soluble we consider it to be a *double* problem: (1) we must establish an appropriate definition for instantaneous velocity and (2) we must find a method for actual evaluation of this quantity. We shall then see that the same argument answers both questions.

We shall later see that this logic is characteristic of many problems in geometry and mechanics which can be solved with the help of mathematical analysis.

Let us now solve our problem. Apart from the instant t for which we want to determine the instantaneous velocity let us consider another later instant $t + \Delta t$. During the time interval Δt which elapses between these two instants the body evidently travels a distance $\Delta s = f(t + \Delta t) - f(t)$ equal to the increment of the function $f(t)$ which corresponds to the increment Δt of the independent variable. We can therefore say that during the time interval Δt which elapses between the instants t and $t + \Delta t$ the average distance covered by the body in unit time (for example, in one second) is equal to:

$$\frac{\Delta s}{\Delta t} = \frac{f(t + \Delta t) - f(t)}{\Delta t}. \quad (1)$$

This relation is known as *average velocity* of the body between the instants t and $t + \Delta t$. But does it tell us anything about the instantaneous velocity of the body at the moment t ? If Δt is large, then during this time interval the velocity of the body can change

many times within wide limits and therefore we are unable to judge the instantaneous velocity of the body at the moment t from the knowledge of its average velocity. But when Δt is small we can assume that during this time interval the velocity of the body will not change greatly and the body will move with approximately the same velocity at different moments of this time interval; therefore the average velocity of the body during this time interval gives a good indication of the approximate instantaneous velocity of the body at the instant t . We can say more accurately that the closer is (1) to the velocity of the body at the instant t in which we are interested, the smaller is the increment Δt ; and even more exactly: we can consider (1) to be *as close as we please* to the required velocity of the body at the instant t provided the time interval Δt is *sufficiently small*. If we denote the required velocity by $v(t)$, then this means: $|\Delta s / \Delta t - v(t)|$ can be as small as we please provided Δt is sufficiently small. This statement can be rephrased in terms of the theory of limits and it is: $v(t)$ is the limit of the relation (1) for $\Delta t \rightarrow 0$:

$$v(t) = \lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}. \quad (2)$$

Hence *instantaneous velocity of a moving body is the limit to which the ratio of the covered path to the elapsed time tends when the latter tends to zero*. We have thus defined the concept of instantaneous velocity and found a method for its evaluation, *i.e.* we have solved both questions of our initial problem.

Note 1. In formula (1) we should consider t (the instant for which we want to establish instantaneous velocity) as a constant; the process which forms the basis of limit transitions implies that the time interval Δt tends to decrease indefinitely while the initial instant of this time interval remains constant. It is evident that the instant t can be chosen arbitrarily but once it is chosen it should remain constant during the process of calculating the velocity.

Note 2. The limit of the relation (2) may or may not exist depending on the choice of the instant t and on the form of the function $f(t)$. If it does not exist, then at the corresponding instant the velocity of the moving body cannot be determined by the method given above. In such cases it is not advisable to seek another definition for the concept of instantaneous velocity but assume simply that at such an instant no instantaneous velocity exists.

Example 1 (uniformly accelerated motion—falling of bodies in vacuum due to gravity). $s = f(t) = gt^2/2$, where g is a constant (the so-called *gravitation constant*) :

$$\Delta s = f(t + \Delta t) - f(t) = \frac{g(t + \Delta t)^2}{2} - \frac{gt^2}{2} = gt\Delta t + \frac{g(\Delta t)^2}{2};$$

$$\frac{\Delta s}{\Delta t} = \frac{f(t + \Delta t) - f(t)}{\Delta t} = gt + \frac{g\Delta t}{2};$$

$$v(t) = \lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t} = gt.$$

Hence the velocity of a freely falling body in vacuum increases in proportion to the elapsed time.

Example 2 (simple harmonic vibrations). $s = f(t) = a \sin \omega t$ (where a and ω are positive constants).

Here s denotes the distance of the moving body from an initial position which is considered to be positive in one direction (for example, to the right) and negative in the other direction (to the left).

$$\begin{aligned} \Delta s &= f(t + \Delta t) - f(t) = a \sin \omega(t + \Delta t) - a \sin \omega t = \\ &= 2a \cos \omega \left(t + \frac{\Delta t}{2} \right) \sin \frac{\omega \Delta t}{2}; \end{aligned}$$

$$\frac{\Delta s}{\Delta t} = a \omega \cos \omega \left(t + \frac{\Delta t}{2} \right) \frac{\sin \frac{\omega \Delta t}{2}}{\frac{\omega \Delta t}{2}};$$

$$v(t) = a \omega \cos \omega t.$$

In this example the velocity of the moving body (if we assume that it moves along a straight line) evidently varies continuously between $s = a$ and $s = -a$ (oscillations with constant amplitude). In accordance with our agreement as to the sign of s we evidently have a positive velocity when the body moves to the right and a negative velocity when it moves in the opposite direction. At the points $s = \pm a$ we should have $\sin \omega t = \pm 1$ and therefore $\cos \omega t = 0$; hence at these points the instantaneous velocity becomes zero; this can readily be understood, for at these points changes in the direction of motion occur and therefore the sign of the velocity also changes. The maximum velocity $|v(t)| = a\omega$ is attained when $\cos \omega t = \pm 1$;

at that instant $x = a \sin \omega t = 0$, *i.e.* the body passes through the initial position.

§ 27. Local density of a heterogeneous rod

A rod is a physical body whose form approaches a section of a straight line; its cross-section is small and constant along its whole length. A rod is said to be *homogeneous* if any two cuts of the same length have the same mass (or, what is the same, the same weight); in a homogeneous rod the masses of any two sections are proportional to their lengths, so that the ratio d of the mass of an arbitrary section to its length is constant and is the same for all sections. The quantity d can be regarded as the mass per unit length of the rod; it is usually known as *density* of a homogeneous rod.

If the rod is heterogeneous, *i.e.* its mass is denser at some places than at other places, then, generally speaking, different masses correspond to two sections of the same length. The ratio of the mass of a section to its length will be different for different sections; it is therefore natural to call this ratio *average density* of the given section of the rod. Since in a given section the density of mass can change considerably many times, therefore, generally speaking, the average density of this section does not tell us anything about the particular density of the mass in the immediate neighbourhood of any point on that section in the same way as in the previous paragraph the average velocity of a car during one hour made it impossible to draw conclusions as to the velocity of the car at the given instant.

Thus if we want to determine density of the substance in the immediate neighbourhood of a particular point on the rod, we encounter difficulties of the same kind as those in § 26 when we were trying to assess instantaneous velocity of a moving body. And since these new difficulties are in all respects similar to the old difficulties, therefore we are quite justified in assuming that we shall be able to solve them by the same methods.

Let us take one end of the rod as the origin 0 and denote the abscissa of any point on the rod with respect to this origin by x . The mass of substance along the section $(0, x)$ is a function of x which increases as x increases; let us denote it by $m = f(x)$. The section of the rod between x and Δx (where Δx is an arbitrary positive number) evidently contains the mass

$$\Delta m = f(x + \Delta x) - f(x);$$

and average density of substance along that section is evidently equal to

$$\frac{\Delta m}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

If the number Δx is large, then density can considerably vary along the section $(x, x + \Delta x)$ and we therefore have no reason to assume that the average density will be indicative of the density of the substance in the immediate neighbourhood of the point x on the rod. On the other hand, if Δx is very small, we can assume that density of the substance will not change considerably over the length Δx so that the average density of substance along the section of length Δx will be close to the required density of the substance in the immediate neighbourhood of the point x . The smaller the number Δx is, the more convincing is this argument; as in the previous paragraph, we therefore conclude that we can take the following quantity as a measure of density of substance in the immediate neighbourhood of the point x on the rod :

$$d(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta m}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

(assuming, of course, that the above limit exists). The quantity $d(x)$ so determined is usually known as the *local* (i.e. at the given spot) *density of the rod* at the point x *). We can obviously regard this local density as a measure of the velocity with which the mass of the rod changes with an increase in length; this point of view connects the problem in this paragraph even more closely with the problem solved in § 26; the remaining difference is, first of all, due to the fact that the factor of time is completely absent in our present problem; if there we were seeking the instantaneous rate of change over the distance covered in *the course of time*, we are now trying to establish the local rate of change of the mass of the rod *as its length increases*. Time has no place in this process. It is thus possible to speak of the rate of change of a function in relation to the independent variable irrespective of the real meaning of these two quantities. This generalisation of the concept of velocity is of utmost importance in mathematical theory; we shall analyse it in detail in the next paragraph.

*) The term "local" (at the given spot) is already found in earlier chapters; it is a property which can vary at different points.

§ 28. Definition of a derivative

Let $y = f(x)$ be an arbitrary function of the independent variable x . If y changes uniformly as x changes (as we know, the necessary condition for this to be so is that $f(x)$ should be a linear function: $f(x) = \alpha x + \beta$), then the rate of change of y with respect to x is equal to a constant number α which in its turn is equal to the ratio of the increment of y to the corresponding increment of x ($\alpha = \Delta y / \Delta x$) and this ratio will always be the same irrespective of the initial value of x and its increment Δx . We have seen all this in § 25 where we said that in the general case, when y changes non-uniformly, the problem of changes of y with respect to x cannot be so easily solved. If we pass from a value x of the independent variable to its new value $x + \Delta x$, then y receives the increment $y = f(x + \Delta x) - f(x)$; the ratio $\Delta y / \Delta x$ will differ for different sections, *i.e.* it will generally depend on the initial value of x and the increment Δx of the independent variable. This ratio describes the *average velocity* of change of y with respect to x along the section $(x, x + \Delta x)$. If we wish to find the *local* velocity of this change, *i.e.* the rate of change of y with respect to x in the neighbourhood of the given value x of the independent variable, then, by repeating word by word the arguments used in the two examples considered above, we evidently conclude that this velocity should be defined as the limit

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (1)$$

of the ratio of increment of the function to the increment of the independent variable when the latter tends to zero. All remarks made in connection with the two examples considered above remain valid in the general case. We can only speak of local velocity when the above limit exists; otherwise there is no velocity. The local velocity is in general different at different points (*i.e.* for different values of the independent variable x); in the expression (1) we must regard the values of x as constant during the limiting process (in this case only Δx changes); however, although this constant value can be chosen arbitrarily, the resulting velocity will be different for every choice (therefore we say that it is "local").

The local velocity as determined by the limiting process (1) can either be positive or negative or zero. It is quite easy to establish the true sign of velocity. In fact, if, for example, the limit of the ratio $\Delta y / \Delta x$ for $x \rightarrow 0$ is positive, then, as we know (theorem 2 § 10), this ratio is also positive provided Δx is sufficiently small;

this means that we should have $\Delta y > 0$ for $\Delta x > 0$ and $\Delta y < 0$ for $\Delta x < 0$; in other words, increments of the function and the independent variable have the same sign in all cases. This means, in its turn, that y increases as x increases (and consequently y decreases together with x). If on the other hand velocity is negative, then an analogous argument shows that y should decrease as x increases and vice versa. Hence the sign of the rate of change determines the *direction* of change of the function (*i.e.* it shows whether the function increases or decreases); the rate of change is in each case determined by the absolute value of this quantity.

Finally we note that in all our examples we have confined ourselves to cases where the increment Δx is positive (although in some examples at the end of § 26 all calculations remain valid and lead to the same results, as can readily be shown, for negative Δx). But in expressions of the type (1) we shall always understand the limiting process in the ordinary sense, *i.e.* we shall require that the limit should exist for both $\Delta x \rightarrow +0$ and $\Delta x \rightarrow -0$ and that these two limits should coincide; only when these conditions are satisfied, we shall regard the local rate of change of the function y with respect to x as existing.

We can thus see that from a purely mathematical point of view the calculation of the rate of change of a function always leads to a definite limiting process. When we are given the function $y = f(x)$ and when we choose the value x of the independent variable, it is necessary in every case to evaluate the limit

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

In cases when this limit exists, it has in general different values for x ; it is therefore a function of x which is usually denoted by y' or $f'(x)$ and is known as the *derivative* of the function $y = f(x)$ with respect to the independent variable x . Thus

$$y' = f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

The derivative of the function $y = f(x)$ with respect to the independent variable x is the limit of the ratio of increment of the function to the increment of the independent variable, provided the latter tends to zero.

The operations of finding the derivative $f'(x)$ of the given function $f(x)$ is known as *differentiation* of the function. In order to

evaluate rate of changes which take place in nature and in technical processes, we must learn to differentiate as many classes of functions as possible.

Differentiation of functions is one of the most important operations of mathematical analysis and we must therefore study it in great detail. The science dealing with the laws of differentiation and properties of derivatives is known as *differential calculus*, and is one of the main branches of mathematical analysis. At first we must learn a series of general rules and special methods of differentiation which will ultimately enable us to find derivatives of a very wide class of functions which also includes all elementary functions. We shall do this in the next paragraph.

§ 29. Laws of differentiation

In this paragraph we shall study the general methods of differentiation and evaluate derivatives of certain functions. We shall thus gradually learn how to find derivatives of a very wide class of functions.

1. Derivative of a constant. *Derivative of a constant is equal to zero.*

In more accurate terms this means that if the function $y = f(x)$ is constant along a certain section which contains the point x , then $y' = f'(x) = 0$. In fact, provided Δx is sufficiently small, we have $f(x + \Delta x) = f(x)$ and therefore $\Delta y = 0$; hence $\Delta y / \Delta x = 0$ for $\Delta x \neq 0$ and therefore $y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = 0$.

2. Derivative of a power. *If $y = x^n$ (where n is a positive integer), then $y' = nx^{n-1}$.*

In fact, Newton's binomial formula gives us :

$$\Delta y = (x + \Delta x)^n - x^n = nx^{n-1}\Delta x + \frac{n(n-1)}{2}x^{n-2}(\Delta x)^2 + \dots + (\Delta x)^n,$$

and therefore for $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}\Delta x + \dots + (\Delta x)^{n-1}.$$

All terms from the second term onwards on the right hand side of this equation contain the factors Δx and therefore tend to zero for $\Delta x \rightarrow 0$. Hence in the limit

$$y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = nx^{n-1}.$$

3. Derivative of a sum. If

$$y = u_1 \pm u_2 \pm \dots \pm u_n,$$

where u_1, u_2, \dots, u_n are functions of x which have a derivative at the point x , we have

$$y' = u'_1 \pm u'_2 \pm \dots \pm u'_n,$$

or in short: *derivative of an algebraic sum is equal to the algebraic sum of the derivatives.* In fact, let us assume that x receives an increment Δx , the functions u_1, u_2, \dots, u_n, y receive corresponding increments $\Delta u_1, \Delta u_2, \dots, \Delta u_n, \Delta y$. It follows from (1) that for the new value $x + \Delta x$ of the independent variable we have the following expression:

$$y + \Delta y = (u_1 + \Delta u_1) \pm (u_2 + \Delta u_2) \pm \dots \pm (u_n + \Delta u_n). \quad (2)$$

Subtracting (1) from (2) we have

$$\Delta y = \Delta u_1 \pm \Delta u_2 \pm \dots \pm \Delta u_n,$$

and therefore for $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u_1}{\Delta x} \pm \frac{\Delta u_2}{\Delta x} \pm \dots \pm \frac{\Delta u_n}{\Delta x};$$

and, finally, taking the limit for $\Delta x \rightarrow 0$ we find that y' exists and that

$$y' = u'_1 \pm u'_2 \pm \dots \pm u'_n.$$

4. A constant factor can be taken outside the symbol of differentiation. More accurately: *if $y = au$, where a is a constant and u a function of x , and if u has a derivative at a certain point, then y' exists at that point and $y' = au'$.* In fact, let us assume that when x receives an increment Δx , u and y receive corresponding increments Δu and Δy . We thus have

$$y + \Delta y = a(u + \Delta u);$$

subtracting the equation $y = au$ term-by-term from this we find:

$$\Delta y = a \Delta u,$$

and therefore for $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = a \frac{\Delta u}{\Delta x}.$$

Finally taking the limit for $\Delta x \rightarrow 0$ we find that y' exists and is equal to $y' = au'$.

5. Derivative of a polynomial. The four laws which we have so far established lead us to a very important result: they show that every polynomial $y = a_0 x^n + a_1 x^{n-1} + \dots + a_n$ has a derivative for every value of x and it enables us to write the expression for this derivative. In fact, applying the rules established above we can readily see that

$$y' = na_0 x^{n-1} + (n-1)a_1 x^{n-2} + \dots + a_{n-1}.$$

Hence *derivative of a polynomial is always a polynomial with one degree less than the degree of the given polynomial.*

6. Derivative of a product. Let $y = uv$, where u and v are functions of x which have derivatives at the point x . Using the usual symbols we have:

$$y + \Delta y = (u + \Delta u)(v + \Delta v),$$

and on subtracting we obtain

$$\Delta y = u \Delta v + v \Delta u + \Delta u \Delta v,$$

and therefore for $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = u \frac{\Delta v}{\Delta x} + v \frac{\Delta u}{\Delta x} + \Delta u \frac{\Delta v}{\Delta x}.$$

When $\Delta x \rightarrow 0$, we should consider u and v to be constant on the right hand side (they depend on x but not on Δx) while Δu and Δv tend to zero (this follows from theorem 8 §11, since the ratios $\Delta u/\Delta x$ and $\Delta v/\Delta x$ have limits in accordance with our assumption). Therefore the limit of the last term on the right hand side is equal to $0.v' = 0$ and limiting process gives:

$$y' = uv' + vu' :$$

derivative of a product of two functions is equal to product of the first factor and the derivative of the second plus product of the second factor and the derivative of the first.

Note. Existence of the derivative y' is at first not assumed but proved in the course of the argument in the same way as was done in deducing laws 3 and 4. The full statement of this rule should therefore be as follows : *if the functions u and v have derivatives at certain points, then the function $y = uv$ also has a derivative at that point, and $y' = uv' + vu'$.* This note applies equally to all subsequent laws of this type.

By using simple induction the reader will be able to extend the above law to include any number of factors.

If $y = u_1 u_2 \dots u_n$, then (provided the derivatives of the functions u_1, u_2, \dots, u_n exist)

$y' = u'_1 u_2 \dots u_n + u_1 u'_2 \dots u_n + \dots + u_1 u_2 \dots u'_n$:
in order to write down derivative of a product containing any desired number of factors we must differentiate one factor and multiply the derivative so obtained by the product of all remaining factors ; we must repeat this process in all possible ways and add all the products so obtained.

We leave to the reader as an exercise to prove that the laws 2 and 4 can be obtained from the above law and that they are, in fact, particular cases of this kind ; the extension of old results on a new basis is always instructive and can often be used as a useful check for the results obtained.

7. Derivative of a quotient. Let $y = u/v$, where u and v are functions of x which have derivatives at the point x , and let $v \neq 0$ at this point. Using the usual symbols we find :

$$y + \Delta y = \frac{u + \Delta u}{v + \Delta v},$$

and by subtracting we obtain :

$$\Delta y = \frac{u + \Delta u}{v + \Delta v} - \frac{u}{v} = \frac{v \Delta u - u \Delta v}{v(v + \Delta v)}.$$

Hence for $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = \frac{v \frac{\Delta u}{\Delta x} - u \frac{\Delta v}{\Delta x}}{v(v + \Delta v)};$$

here, as in the above deduction, u and v are constant for $\Delta x \rightarrow 0$ and $\Delta v \rightarrow 0$; therefore the limiting process proves the existence of y' and gives us the following expression for the derivative

$$y' = \frac{vu' - uv'}{v^2}. \quad (3)$$

In particular, when the functions u and v are polynomials, the ratio u/v represents a rational fraction; formula (3) thus shows that derivative of a rational fraction is always a rational fraction.

8. Derivatives of trigonometrical functions. (a) Let $y = \sin x$; then

$$y + \Delta y = \sin(x + \Delta x),$$

$$\Delta y = \sin(x + \Delta x) - \sin x = 2 \cos\left(x + \frac{\Delta x}{2}\right) \sin \frac{\Delta x}{2},$$

$$\frac{\Delta y}{\Delta x} = 2 \cos\left(x + \frac{\Delta x}{2}\right) \frac{\sin \frac{\Delta x}{2}}{\Delta x} = \cos\left(x + \frac{\Delta x}{2}\right) \frac{\sin \frac{\Delta x}{2}}{\frac{\Delta x}{2}};$$

the last factor tends to unity as its limit for $\Delta x \rightarrow 0$; on the other hand it follows from continuity of the function $\cos x$ (cf. § 24) that

$$\cos\left(x + \frac{\Delta x}{2}\right) \rightarrow \cos x \quad (\Delta x \rightarrow 0),$$

therefore the limiting process proves the existence of the limit

$$\lim_{\Delta x \rightarrow 0} \left(\frac{\Delta y}{\Delta x}\right) = y' \text{ and gives:}$$

$$y' = \cos x.$$

(b) Let $y = \cos x$; a similar argument, which we leave to the reader, gives

$$y' = -\sin x.$$

(c) Let $y = \tan x = \sin x / \cos x$; we are dealing here with the ratio of two functions whose derivatives are already found; assuming that $\sin x = u$, $\cos x = v$, we find, according to formula (3):

$$y' = \frac{\cos x \cos x - \sin x (-\sin x)}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

(d) If $y = \cot x = \cos x / \sin x$, then a similar calculation gives

$$y' = -\frac{1}{\sin^2 x}.$$

The reader will have no difficulty in finding derivatives of the functions $y = \sec x = 1/\cos x$ and $y = \operatorname{cosec} x = 1/\sin x$.

9. Before going further we must introduce another very important limit relationship. In § 17 we have shown that the expression $(1 + 1/n)^n$ tends to a definite limit when n runs successively through the series of natural numbers, and we denoted this limit by e ; let us now assume that in the expression $(1 + 1/x)^x$ the variable x increases indefinitely ($x \rightarrow +\infty$) and runs through all intermediate values; we can see that we have in this case:

$$\lim \left(1 + \frac{1}{x}\right)^x = e. \quad (4)$$

In fact, let us denote by n the greatest integer for a given value of x which does not exceed x so that

$$n \leq x < n + 1.$$

Therefore we evidently have for $x \geq 1$

$$\left(1 + \frac{1}{n+1}\right)^n < \left(1 + \frac{1}{x}\right)^x < \left(1 + \frac{1}{n}\right)^{n+1},$$

or

$$\frac{\left(1 + \frac{1}{n+1}\right)^{n+1}}{1 + \frac{1}{n+1}} < \left(1 + \frac{1}{x}\right)^x < \left(1 + \frac{1}{n}\right)^n \left(1 \pm \frac{1}{n}\right);$$

if $x \rightarrow +\infty$, then evidently $n \rightarrow +\infty$; on the left-hand side of the above inequalities the numerator tends to the limit e and the denominator tends to unity, while on the right-hand side the first factor tends to e and the second factor tends to unity. Thus the left and right-hand sides tend to the same limit e for $x \rightarrow +\infty$; hence the main parts of the inequalities also tend to the same limit, which proves the relation (4).

Let us now assume that $x \rightarrow -\infty$ and $y = -x$, so that $y \rightarrow +\infty$ and

$$\begin{aligned} \left(1 + \frac{1}{x}\right)^x &= \left(1 - \frac{1}{y}\right)^{-y} = \left(\frac{y}{y-1}\right)^y = \left(1 + \frac{1}{y-1}\right)^y \\ &= \left(1 + \frac{1}{y-1}\right)^{y-1} \left(1 + \frac{1}{y-1}\right); \end{aligned}$$

the first term on the right-hand side tends to the limit e for $y \rightarrow +\infty$ in accordance with the proved proposition (since $y - 1 \rightarrow +\infty$) and the second term evidently tends to unity; this shows that

$$\left(1 + \frac{1}{x}\right)^x \rightarrow e \quad (x \rightarrow -\infty),$$

i. e. that the relation (4) applies every time for $|x| \rightarrow +\infty$ irrespective of the sign of x .

Let us now assume that in the expression

$$(1 + \alpha)^{\frac{1}{\alpha}}$$

α tends to zero in an arbitrary way (we are only assuming that $\alpha \neq 0$, since the above expression becomes void); assuming that $1/\alpha = x$ we have :

$$(1 + \alpha)^{\frac{1}{\alpha}} = \left(1 + \frac{1}{x}\right)^x;$$

we have $|x| \rightarrow +\infty$ for $\alpha \rightarrow 0$ and therefore in accordance with the above proof $(1 + 1/x)^x \rightarrow e$; the above equation therefore shows that

$$\lim_{\alpha \rightarrow 0} (1 + \alpha)^{\frac{1}{\alpha}} = e. \quad (5)$$

We have thus proved the above relation and we can therefore use it in future.

10. Derivative of a logarithm. Let $y = \log_a x$, where $a \neq 1$ is a constant positive number and $x > 0$. In this case

$$y + \Delta y = \log_a (x + \Delta x),$$

$$\Delta y = \log_a (x + \Delta x) - \log_a x = \log_a \left(1 + \frac{\Delta x}{x}\right),$$

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x} \log_a \left(1 + \frac{\Delta x}{x}\right) = \frac{1}{x} \frac{x}{\Delta x} \log_a \left(1 + \frac{\Delta x}{x}\right) = \frac{1}{x} \log_a \left\{\left(1 + \frac{\Delta x}{x}\right)^{\frac{x}{\Delta x}}\right\}.$$

Let us now assume that $\Delta x/x = \alpha$; hence we have $\alpha \rightarrow 0$ for $\Delta x \rightarrow 0$ and it therefore follows from the relation (5) that

$$\left(1 + \frac{\Delta x}{x}\right)^{\frac{x}{\Delta x}} = (1 + \alpha)^{\frac{1}{\alpha}} \rightarrow e;$$

and since the function $\log_a x$ is a continuous function,

$$\log_a \left\{ \left(1 + \frac{\Delta x}{x} \right)^{\frac{x}{\Delta x}} \right\} \rightarrow \log_a e \quad (\Delta x \rightarrow 0);$$

hence $\lim_{\Delta x \rightarrow 0} \left(\frac{\Delta y}{\Delta x} \right) = y'$ exists and

$$y' = \frac{1}{x} \log_a e.$$

This result is remarkable insofar as the derivative of a transcendental function $\log_a x$ is a simple rational function of the form c/x , where c is a constant. The form of this derivative will be particularly simple if we choose e as the base of the logarithmic system, for in that case $\log_e e = \log_e e = 1$ and

$$y' = \frac{1}{x}.$$

We shall later see that many other analytical formulae are obtained in a particularly simple form when e is taken as the logarithmic base. For this reason e is usually taken as the logarithmic base in analysis; logarithms with the base e are known as “natural”; the natural logarithm of x is denoted by the symbol $\ln x$: thus if $y = \ln x$, then $y' = 1/x$; if, on the other hand $y = \log_a x$, then, as we have seen,

$$y' = \frac{1}{x} \log_a e;$$

but when taking the logarithm with the base a of $a = e^{\ln a}$ we obtain:

$$\log_a a = 1 = \ln a \log_a e, \quad \log_a e = \frac{1}{\ln a},$$

so that we can write down the above equation in the form

$$y' = \frac{1}{x \ln a}.$$

11. Derivative of a composite function. Let y be a composite function of x i.e. y is given as a function of an intermediate function u , $y = f(u)$, and u as a function of x , $u = \varphi(x)$; so that

$$y = f[\varphi(x)]; \quad (6)$$

we must find derivative of the function (6) (*i. e.* differentiate y with respect to x) while knowing the derivatives of the functions $f(u)$ and $\varphi(x)$ (*i. e.* when we are able to differentiate y with respect to u and u with respect to x).

Let us assume that x receives the increment Δx ; in that case u receives a corresponding increment Δu and therefore y receives the increment Δy ; if $\Delta x \rightarrow 0$, then $\Delta u \rightarrow 0$, $\Delta y \rightarrow 0$, assuming that

$$\alpha = \begin{cases} \frac{\Delta y}{\Delta u} - f'(u), & \text{if } \Delta u \neq 0, \\ 0, & \text{if } \Delta u = 0. \end{cases}$$

Owing to the fact that we have $\Delta y / \Delta u \rightarrow f'(u)$ for $\Delta u \rightarrow 0$, therefore evidently $\alpha \rightarrow 0$ for $\Delta x \rightarrow 0$; further, for $\Delta u \neq 0$ it follows from the definition of α that :

$$\Delta y = f'(u) \Delta u + \alpha \Delta u$$

but it is clear that this relation is also valid for $\Delta u = 0$, and therefore it is always valid; dividing both sides of this relation by Δx we have :

$$\frac{\Delta y}{\Delta x} = f'(u) \frac{\Delta u}{\Delta x} + \alpha \frac{\Delta u}{\Delta x};$$

but we have $\Delta u / \Delta x \rightarrow \varphi'(x)$ and $\alpha \rightarrow 0$ for $\Delta x \rightarrow 0$; therefore the limiting process proves the existence of the limit $\lim_{\Delta x \rightarrow 0} \left(\frac{\Delta y}{\Delta x} \right) = y'$ and gives

$$y' = f'(u) \varphi'(x) = f'[\varphi(x)] \varphi'(x). \quad (7)$$

We can therefore see that *derivative of a composite function is equal to product of the derivative of the given function with respect to the intermediate variable and the derivative of the intermediate variable with respect to the independent variable*. Hence in order to find derivative of a composite function which is given in the form of a two-link chain $y = f(u)$ and $u = \varphi(x)$, we must simply differentiate each link of the chain separately and multiply the derivatives so obtained.

Example 1. $y = \sin kx$, where k is a constant; let us assume that $kx = u$ so that

$$y = \sin u, \quad u = kx;$$

it follows from formula (7) that

$$y' = \cos u \cdot k = k \cos kx.$$

Example 2. $y = \ln \cos x$, $\cos x = u$, $y = \ln u$; according to formula (7) we have

$$y' = \frac{1}{u} (-\sin x) = -\frac{\sin x}{\cos x} = -\tan x.$$

By using simple induction we can extend the above law for differentiation of composite function to include functions given in the form of a chain of three or more links :
thus, if

$$y = f(u), \quad u = \varphi(v), \quad v = \psi(x),$$

then derivative of y with respect to x can be found by the formula

$$y' = f'(u) \varphi'(v) \psi'(x) + f' \{ \varphi [\psi (x)] \} \varphi' [\psi (x)] \psi' (x).$$

12. Derivative of an inverse function. We know that the relation which defines y as a function of x enables us in certain cases to define the inverse function, i.e. x as a function of y . Let $y = f(x)$ and let the inverse function be $x = \varphi(y)$. Let us assume that the nonzero derivative $f'(x)$ exists at a point $x = \varphi(y)$; the increment Δy of y corresponds to the increment Δx of x ; since $\Delta y = f(x + \Delta x) - f(x)$, therefore we must have $\Delta x \neq 0$ for $\Delta y \neq 0$; thus for $\Delta y \neq 0$

$$\frac{\Delta x}{\Delta y} = \frac{1}{\frac{\Delta y}{\Delta x}}. \quad (8)$$

Let us now assume that $\Delta y \rightarrow 0$; if the function $x = \varphi(y)$ is continuous at the point y , we have $\Delta x \rightarrow 0$ and therefore

$$\frac{\Delta y}{\Delta x} \rightarrow f'(x) \neq 0;$$

the relation (8) thus shows that the ratio $\Delta x / \Delta y$ tends to $1/f'(x)$ for $\Delta y \rightarrow 0$; in other words, the derivative $\varphi'(y)$ exists and is equal to $1/f'(x)$. We thus arrive at the following law for differentiating an inverse function :

If the function $y = f(x)$ has a nonzero derivative at the point x and the inverse function $x = \varphi(y)$ is continuous at the point y , then $\varphi'(y)$ exists and is equal to $1/f'(x)$.

13. Derivatives of exponential functions. Let $y = a^x$, where a is a constant positive number; in that case $x = \log_a y$; according to 10 the derivative of x with respect to y is equal to

$$x' = \frac{1}{y \ln a};$$

and it follows from the above law for differentiating inverse functions that

$$y' = \frac{1}{x'} = y \ln a = a^x \ln a;$$

in particular, if $y = e^x$, then $y' = e^x$, i.e. the “simple” exponential function is invariant in differentiation: derivative of this function is equal to the function itself.

If $y = e^{\alpha x}$, where α is a constant, then we can regard y as a composite function of x by assuming that $\alpha x = u$; it readily follows from (7) that

$$y' = \alpha e^{\alpha x} = \alpha y.$$

In practice one often comes across the so-called “hyperbolic functions”, e.g. the “hyperbolic cosine”

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

and the “hyperbolic sine”

$$\sinh x = \frac{e^x - e^{-x}}{2};$$

the reader will be able to show by himself that each of these two functions is the derivative of the other.

14. Derivative of a power function. Let $y = x^\alpha$; where α is an arbitrary constant. We have seen in 2 that if α is a natural number, then

$$y' = \alpha x^{\alpha-1};$$

we will now show that this formula remains valid for every α .

We can write

$$y = x^\alpha = e^{\alpha \ln x};$$

and assuming that $\alpha \ln x = u$, we have

$$y = e^u, \quad u = \alpha \ln x,$$

and according to the law for differentiating composite functions

$$y' = e^u \cdot \frac{\alpha}{x} = x^\alpha \cdot \frac{\alpha}{x} = \alpha x^{\alpha-1},$$

which we wanted to prove.

In a particular case when $\alpha = \frac{1}{2}$ we have :

$$y = \sqrt{x}, \quad y' = \frac{1}{2\sqrt{x}}.$$

When $\alpha = -1$, $y = 1/x$, $y' = -1/x^2$, and so on.

This method can be used for finding derivatives of a much wider class of functions

$$y = \{f(x)\}^{\varphi(x)},$$

where $f(x)$ and $\varphi(x)$ are differentiable functions. In fact,

$$y = e^{\varphi(x) \ln f(x)} = e^u, \quad u = \varphi(x) \ln f(x),$$

and therefore in accordance with the law for differentiating composite functions

$$y' = e^u \{\varphi(x) \ln f(x)\}' = \{f(x)\}^{\varphi(x)} \left\{ \varphi(x) \frac{f'(x)}{f(x)} + \varphi'(x) \ln f(x) \right\};$$

Example. $y = x^x$, $y' = x^x \{1 + \ln x\}$.

15. Derivatives of inverse trigonometrical functions.

(a) Let $y = \arcsin x$ for $-1 < x < 1$ so that y increases from $-\pi/2$ to $+\pi/2$ as x changes along the section $(-1, +1)$. Since in this case $x = \sin y$, therefore, as a result of the law established in 11,

$$y' = \frac{1}{x'} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}} = \frac{1}{\sqrt{1 - x^2}},$$

where the positive square root should be taken so that $\cos y > 0$ for $-\pi/2 < y < +\pi/2$. Hence for $y = \arcsin x$ we have

$$y' = \frac{1}{\sqrt{1 - x^2}} \quad (-1 < x < 1).$$

We can similarly prove that

(b) if $y = \arccos x$, then

$$y' = -\frac{1}{\sqrt{1 - x^2}} \quad (-1 < x < 1);$$

(c) if $y = \arctan x$, then

$$y' = \frac{1}{1 + x^2} \quad (-\infty < x < +\infty);$$

(d) if $y = \operatorname{arccot} x$, then

$$y' = -\frac{1}{1+x^2} \quad (-\infty < x < +\infty).$$

Note 1. It is interesting to note that inverse trigonometrical functions which are transcendental have very simple derivatives expressed in terms of very simple algebraic functions (in the case of $\operatorname{arccot} x$ and $\operatorname{arctan} x$ these functions are even rational) (we have an analogous case in differentiating logarithms).

Note 2. It is also interesting to note that the derivatives of $\arcsin x$ and $\arccos x$ differ from one another only in sign (the same also applies to the derivatives of $\operatorname{arctan} x$ and $\operatorname{arccot} x$); this can easily be foreseen if the following wellknown trigonometrical identities are differentiated :

$$\arcsin x + \arccos x = \frac{\pi}{2}, \quad \operatorname{arctan} x + \operatorname{arccot} x = \frac{\pi}{2}.$$

16. In this chapter we have learnt to differentiate all simple elementary functions; derivatives of these functions are listed in the table below :

y	y'	y	y'	y	y'	y	y'
a	0	a^x	$a^x \ln a$	$\tan x$	$\frac{1}{\cos^2 x}$	$\operatorname{arctan} x$	$\frac{1}{1+x^2}$
x^α	$\alpha x^{\alpha-1}$	$\ln x$	$\frac{1}{x}$	$\cot x$	$-\frac{1}{\sin^2 x}$	$\operatorname{arccot} x$	$-\frac{1}{1+x^2}$
$\frac{1}{x}$	$-\frac{1}{x^2}$	$\log_a x$	$\frac{1}{x \ln a}$	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$	$\sinh x$	$\cosh x$
\sqrt{x}	$\frac{1}{2\sqrt{x}}$	$\sin x$	$\cos x$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$	$\cosh x$	$\sinh x$
e^x	e^x	$\cos x$	$-\sin x$				

The laws of differentiation which we have established above enable us to find without difficulty derivatives of any combination of functions obtained as a result of algebraic operations or by constructing any number of composite functions. In order to make the reader appreciate how wide this class of functions really is, which he is now able to differentiate, we recommend him to try, as an exercise, to find a function for which he cannot find a derivative. The difficulty of this problem will convince him that his ability to

differentiate is very wide indeed. However, the knowledge of principles is insufficient; every mathematician should learn to differentiate quickly and unmistakably and in order to learn this he must do many exercises.

Many exercises for differentiation are given in the problem book by B.P. Demidovich, Section II, § 1. In addition to these exercises (problems 14-68) we would also recommend the student to think over some other problems (for example, problems 72, 73, 79, 90, 21).

§ 30. Existence of functions and their geometrical illustrations

The given function $y = f(x)$ has a derivative at the point x if and only if the following limit exists :

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = y'.$$

It can readily be seen that it is necessary for this limit to exist that the function $f(x)$ should be continuous at the point x ; in fact, it follows from theorem 8 § 11 that the ratio $\Delta y / \Delta x$ will have a limit for $\Delta x \rightarrow 0$ provided the increment Δy is infinitely small for $\Delta x \rightarrow 0$, and this means that the function $f(x)$ is continuous at the point x . Hence the function $f(x)$ cannot have a derivative at a point of discontinuity; in particular, a function which is everywhere discontinuous cannot have a derivative (let us remind the function $D(x)$ §§ 4, 20).

Can a continuous function have no derivative? It can readily be shown that this is possible. In fact, it may happen (and it frequently does happen) that the limits

$$\lim_{\Delta x \rightarrow +0} \frac{\Delta y}{\Delta x}, \quad \lim_{\Delta x \rightarrow -0} \frac{\Delta y}{\Delta x} \quad (1)$$

exist but do not coincide; a single limit for $\Delta x \rightarrow 0$ does not exist in this case and therefore a derivative does not exist either. This is the case, for example, with the function $y = |x|$ for $x = 0$: since at this point $y = 0$, so Δy coincides with y (and Δx coincides with x); we therefore have:

$$\frac{\Delta y}{\Delta x} = \frac{y}{x} = \frac{|x|}{x};$$

this ratio is equal to unity for every $x > 0$ and it is equal to -1 for every $x < 0$; therefore

$$\lim_{\Delta x \rightarrow +0} \frac{\Delta y}{\Delta x} = 1, \quad \lim_{\Delta x \rightarrow -0} \frac{\Delta y}{\Delta x} = -1.$$

The question also arises whether it may happen that the function $y = f(x)$, which is continuous at the point x , tends to neither of the limits (1) so that there is no left or right derivative? This case is also possible but here the construction of an example is somewhat more difficult. Let us consider the following function which is defined for all values of x :

$$y = f(x) = \begin{cases} x \sin \frac{1}{x} & (x \neq 0), \\ 0 & (x = 0). \end{cases}$$

We have for $x \neq 0$

$$|f(x)| = |x \sin \frac{1}{x}| \leq |x|,$$

therefore $f(x) \rightarrow 0$ for $x \rightarrow 0$ and since $f(0) = 0$, the function $f(x)$ is continuous for $x = 0$. As in the preceding example we have here $\Delta x = x$, $\Delta y = y$; therefore for $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = \frac{y}{x} = \sin \frac{1}{x} = \sin \frac{1}{\Delta x};$$

if n is an arbitrary natural number, then for

$$\Delta x = \frac{2}{(4n+1)\pi} \quad (2)$$

we have

$$\frac{1}{\Delta x} = 2\pi n + \frac{\pi}{2}, \quad \sin \frac{1}{\Delta x} = 1,$$

and for

$$\Delta x = \frac{2}{(4n-1)\pi} \quad (3)$$

$$\frac{1}{\Delta x} = 2\pi n - \frac{\pi}{2}, \quad \sin \frac{1}{\Delta x} = -1.$$

But Δx will pass an infinite number of times through values of the form (2) and (3) as n increases indefinitely for $\Delta x \rightarrow +0$;

$$\frac{\Delta y}{\Delta x} = \sin \frac{1}{\Delta x}$$

varies an infinite number of times between $+1$ and -1 and cannot tend to a limit. This means that the first of the limits (1) does not

exist in this case ; absence of the second limit is proved in exactly the same way.

In all cases considered so far the given function had no derivative at one point only (for one value of x) but derivatives exist at all other points. On the basis of the above examples it is quite easy to construct a continuous function with two, three or an arbitrary number of points where no derivative exists. However, it was believed for a long time that a continuous function should nevertheless have derivatives everywhere except at certain isolated points; this was confirmed in the first place by the geometrical picture which we shall now study; only in the second half of the last century an example of a continuous function was published which had no derivatives at all. At present many methods for constructing such functions are known; they are, however, all too complicated to be given here.

The geometrical representation of a function is, as we know, a very valuable method of investigation, because many characteristics of the function and its behaviour would be difficult to elucidate from its formula (or from a table), while the graph illustrates them quite clearly. Every characteristic of a function should appear as a geometrical property on the representative curve in its graphical representation. It may be foreseen that the graph which represents the function can also give us a visual representation of its derivative. This geometrical analysis of the derivative is very important in analysis as well as in geometry and we shall now proceed with its study

Let us assume that the graph represents the function $y = f(x)$ on the system of cartesian coordinates (x, y) (Fig. 12). Mark the points $M(x, y)$ and $N(x + \Delta x, y + \Delta y)$ on the curve. Draw the line MP parallel to the OX axis. It is evident that in the right-angled triangle MNP the sides adjacent to the right-angle are $MP = \Delta x$ and $NP = \Delta y$. Therefore the ratio $\Delta y / \Delta x$ is equal to the tangent of the angle formed by the chord MN and the positive direction of the OX -axis.

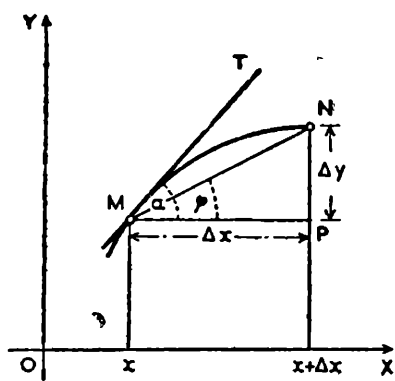


Fig. 12.

Let us now assume that Δx tend to zero. In this case the point M remains stationary while the point N

approaches it indefinitely closely. The chord MN will change its direction and at every moment of this process the gradient of the chord will be :

$$\tan \varphi = \frac{\Delta y}{\Delta x} ;$$

if the given function has a derivative at the point x , *i.e.* if the following limit exists :

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = f'(x) = y',$$

then this means geometrically that the direction of the chord MN tends in this process to a limiting position MT which makes an angle α with the positive direction of the OX -axis, where

$$\tan \alpha = \lim_{\Delta x \rightarrow 0} \tan \varphi = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = y'. \quad (4)$$

The straight line MT can be defined purely geometrically as the limit in a position of the chord MN which joins the point M with the indefinitely approaching point N on the same curve and is known as the *tangent* to the given curve at the point M . Equation (4) shows that *the derivative of the function $f(x)$ at the point x is equal to the gradient of the tangent to the corresponding curve at a point with the abscissa x* . If we assume as usual (in accordance with our visual representation) that the direction of the tangent is characteristic of the direction of the curve itself at the given point, we can directly see that if the curve (as x increases, *i.e.* from left to right) rises, then its derivative is not negative and the steeper the gradient, the greater the derivative; on the other hand, if the curve goes downwards (from left to right), the derivative is not positive and in this case the absolute value of the derivative is the greater, the steeper the gradient is. This geometrical representation is in full agreement with the definition given at the beginning of this chapter where a derivative is defined as the rate of change of y with respect to x ; the quicker y increases as x increases, the steeper the gradient of the curve $y = f(x)$ and the greater therefore the rate y' of this increment.

The above geometrical method of representation of a derivative enables us to understand more clearly the cases where no derivatives exist which we have considered at the beginning of this paragraph. Fig. 13 represents the graph of the function $y = |x|$ and Fig. 14 the

function $y = x \sin(1/x)$. In the first case the line $y = |x|$ has a definite direction to the left and to the right for $x = 0$, but these

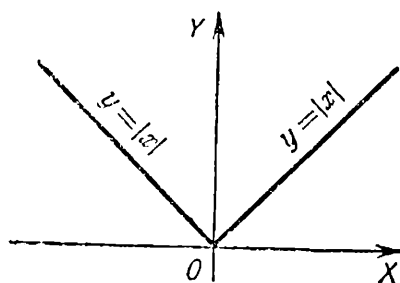


Fig. 13.

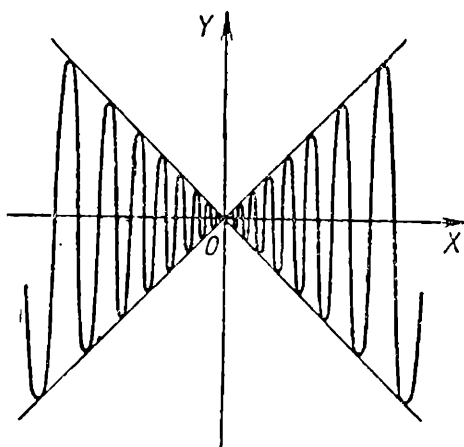


Fig. 14.

directions are not the same; in the second case the curve $y = x \sin(1/x)$ has no definite direction to the left or to the right for $x = 0$ (there is no tangent); as $|x|$ gets smaller and smaller, the direction of the tangent varies repeatedly between the straight lines $y = x$ and $y = -x$ and therefore cannot tend to a limiting direction.

Finally from the point of view of the geometrical interpretation it is easy to understand why it was thought for so long that every continuous function should have a derivative (with the exception of some singularities): in fact, it is very difficult to imagine a continuous curve which would not have a tangent at a single point; and even now when existence of such curves has been established beyond doubt we can imagine such a curve only very approximately; the position of such a curve with respect to its every point is approximately the same as the position of the curve represented in Fig. 14 in the neighbourhood of the point O . However, such curves exist and their discovery was one of the most vivid examples in the history of mathematics that intuition, which reigned for centuries, may sometimes be mistaken.

Finally, let us note that knowledge of the value of the derivative y' makes it possible to use elementary methods for construction of the tangent to the curve $y = f(x)$ at the point M . In elementary geometry we have learnt how to construct a tangent to a circle and in analytical geometry how to construct tangents to all curves of the second order; but only differential calculus shows us how to construct, in general, a tangent to an arbitrary curve at a given point.

CHAPTER VII

DIFFERENTIALS

§ 31. Definition and relationship with derivatives.

If the function $y = f(x)$ has a derivative at the point x ,

$$y' = f'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x},$$

then the quantity

$$\frac{\Delta y}{\Delta x} - y' = \alpha$$

is an infinitesimal for $\Delta x \rightarrow 0$. It therefore follows that

$$\Delta y - y' \Delta x = \alpha \Delta x$$

is an infinitesimal of a higher order as compared to Δx ; using the symbols introduced in § 12 we can denote this quantity by $o(\Delta x)$; hence

$$\Delta y = y' \Delta x + o(\Delta x). \quad (1)$$

Owing to the fact that $y' = f'(x)$ depends only on x and remains constant for $\Delta x \rightarrow 0$, therefore, $y' \Delta x$ is proportional to Δx ; hence the relation (1) shows that *the increment of a function which has a derivative at the point x can be represented as sum of a quantity proportional to Δx and an infinitesimal of a higher order as compared to Δx .*

Conversely : *if the increment of the function $y = f(x)$ at the point x can be represented in the form*

$$\Delta y = a \Delta x + o(\Delta x), \quad (2)$$

where a is independent of Δx , then the function y is differentiable at the point x and $f'(x) = a$. In fact, it follows from (2) that

$$\frac{\Delta y}{\Delta x} = a + o(1),$$

and therefore

$$\frac{\Delta y}{\Delta x} \rightarrow a \quad (\Delta x \rightarrow 0),$$

i.e. $y' = a$.

Example 1. $f(x) = \ln x$; it follows from formula (1) that $f(1+y) - f(1) = f'(1)y + o(y)$ ($y \rightarrow 0$); but $f(1) = 0, f'(1) = 1$ and we have

$$\ln(1+y) = y + o(y),$$

or, which is the same,

$$\ln(1+y) \sim y \quad (y \rightarrow 0);$$

only *natural* logarithms possess this most important property which makes their use so convenient in mathematical analysis.

Example 2. $f(x) = e^x$; it follows from formula (1) that $f(x) - f(0) = f'(0)x + o(x)$ ($x \rightarrow 0$); but $f(0) = f'(0) = 1$ and therefore

$$e^x - 1 = x + o(x),$$

or, which is the same,

$$e^x - 1 \sim x \quad (x \rightarrow 0).$$

The expression for Δy given in formula (1) is exceptionally important, for it shows that the increment of the function given accurately up to infinitesimals of higher orders can be represented by a linear function of the increment of the independent variable. In this formula the first term $y' \Delta x$ which is proportional to Δx is said to be *differential* of the function y and is denoted by dy so that

$$\Delta y = dy + o(\Delta x). \quad (3)$$

Hence *differential of a function is product of the derivative of this function and the increment of the independent variable*, so that for example

$$d \sin x = \cos x \Delta x,$$

$$d \ln x = \frac{\Delta x}{x},$$

etc. It means that in order to define differential of a function it is necessary to know the initial value of the independent variable x and its increment Δx ; only thereafter it is possible to define fully the differential of a function and evaluate it.

We have seen above that if the increment Δy of the function y can be represented in the form (2), then the first term on the right-hand side is equal to $y' \Delta x = dy$; therefore differential of the function at a given point can simply be defined as a quantity proportional to Δx which differs from Δy by an infinitesimal of a higher order as compared to Δx . Such a quantity is often called *principal linear part* of the increment Δy . We can therefore say that *differential of a function* (for given x and Δx) *is principal linear part of its increment*. We can also see that *in order that a function should be differentiable at the given point it is necessary and sufficient that its increment should have a principal linear part*.

This definition defining a differential as principal linear part of the increment is very important, for it serves as the basis of the most important application of differentials. We shall later see that existence of derivatives and existence of principal linear part of the increment are no longer equivalent requirements for functions of several variables; it is noteworthy that the most natural definition of differentiability of a function in such cases is, as we shall see later, not the existence of derivative but the existence of principal linear part of the increment.

The theoretical and the immediate practical (calculative) significance of a differential is mostly based on formula (3). The dependence of Δy on Δx is generally rather complicated and calculation of the accurate value of Δy for given x and Δx is rather difficult. The relation (3) however, shows, that if Δx is small, the approximate evaluation of Δy can be successfully replaced by evaluation of dy , for the difference between these quantities (i.e. the error due to this replacement) is an infinitesimal of a higher order as compared to Δx and therefore comprises only of a negligible part of the evaluated quantity for small Δx (provided, of course, that $y' \neq 0$). As a rule it is always much simpler to evaluate dy than Δy , for the dependence of dy on Δx is linear.

Let us now consider a simple example. Let us assume that we want to evaluate approximately the expression $\ln(2 + \alpha)$, where α is very small.

The differential of the function $\ln x$ is equal to $\Delta x/x$; when $x = 2$, this differential is equal to $\Delta x/2$; therefore assuming that $\Delta x = \alpha$ we find from formula (3)

$$\ln(2 + \alpha) - \ln 2 = \frac{\alpha}{2} + o(\alpha),$$

and therefore

$$\ln(2 + \alpha) = \ln 2 + \frac{\alpha}{2} + o(\alpha);$$

thus by knowing $\ln 2$ we can immediately find the value of $\ln(2 + \alpha)$ with a good approximation for sufficiently small values of α ; thus

$$\ln 2.001 \approx \ln 2 + 0.0005,$$

$$\ln 2.002 \approx \ln 2 + 0.001,$$

$$\ln 2.003 \approx \ln 2 + 0.0015,$$

etc. It is clear that this method is very useful, for example, for compilation of logarithmic tables. Obviously in every case an assessment of the error $o(\Delta x)$ incurred as a result of replacement of the increment of the function Δy by its differential dy must be given. This assessment necessitates further development of the theory and we shall find later on that it can be found. The student will find useful exercises in the problem book by B. P. Demidovich, Section II, Nos. 144, 145, 159, 160, 164.

Since derivative of the function $y = x$ is equal to unity for every value of x , therefore differential of this function is simply equal to Δx for every value of x so that the increment and differential of the function $y = x$ coincide*):

$$\Delta x = dx;$$

we can therefore replace Δx by dx in the expression for differential of an arbitrary function $y = f(x)$; this gives

$$dy = y' dx,$$

and therefore

$$y' = \frac{dy}{dx}; \quad (4)$$

the derivative is equal to the ratio of differential of the function to differential of the independent variable. The expression (4) for the derivative is very convenient, for its symbols are used as much as the symbols y' and $f'(x)$; it is, of course, somewhat more complicated but its advantage is due to the fact that it clearly shows the variable x with respect to which we differentiate. This is particularly important in cases where the problems includes derivatives of one function with

*) This evidently also applies to every linear function $y = \alpha x + \beta$.

respect to different variables. Thus when we differentiate a composite function given by a two-link chain $y = f(u)$, $u = \varphi(x)$ (§ 29), we must deal with derivatives of y with respect to the independent variable x and to the intermediate function u ; the notation y' is here less convenient, for it does not directly show which of the derivatives in question it symbolises; on the other hand, by using the notation (4) we write in such cases dy/dz and dy/du respectively and directly see with respect to which variable we differentiate (the notation dy/du requires further explanations which will be given in § 33).

The relation (4) is of utmost importance in further development of differential calculus as we shall see in the following paragraphs.

§ 32. Geometrical illustration and laws for evaluation.

Like any other quantity which is determined by the course of the function $y = f(x)$ the graphical representation of differential of this function should present an appropriate geometrical picture. Fig. 15 represents a detail of Fig. 12. Here MT represents the tangent to the curve $y = f(x)$ at the point M with co-

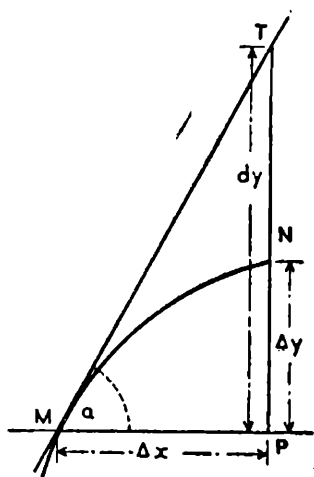


Fig. 15.

ordinates (x, y) . In the right-angled triangle MTP the side TP adjacent to the right angle is equal to the side MP , which is also adjacent to the right angle, multiplied by the tangent of the angle α ; but $MP = \Delta x$, $\tan \alpha = y' = f'(x)$; therefore

$$TP = y' \Delta x = f'(x) \Delta x = dy;$$

thus in our diagram the differential of the function $y = f(x)$ which corresponds to the given values of x and Δx is represented by TP which is evidently equal to the increment of the ordinate of the tangent MT from x to $x + \Delta x$ (at the same time the increment of the ordinate of the curve $y = f(x)$ itself along the same path). Since $\Delta x = dx$, so the relation (4) §31 of the derivative and the differential is given in fig. 15 by an elementary trigonometrical formula

$$\tan \alpha = \frac{TP}{MP},$$

which connects the sides adjacent to the right angle in a right-angled triangle with one of its acute angles.

The *mechanical* illustration of a differential is also interesting. If $s = f(t)$ is the law of motion of a body, then, as we know (c.f. §26), $s' = f'(t)$ represents the instantaneous velocity of this motion at the instant t . The differential of the path

$$ds = s' \Delta t = f'(t) \Delta t$$

is therefore equal to the length of the path covered by the body if at the instant Δt it is moving with the same velocity as at the moment t (i.e. if its velocity from the instant t onwards remains unchanged during the time interval Δt). When we say that the car moves at a given instant with the velocity of 40 km per hour, we really mean that it has covered a distance of 40 km during the past hour provided it constantly maintained the same speed as at the given instant. Hence this number (40 km) represents the differential of the path covered by the car at the given instant (when $\Delta t = 1$ hour).

Finding differential of a given function, like finding its derivative, is known as *differentiation* of this function; the fact, that these two operations are known by the same name is natural and intelligible: if the derivative y' is known, then in order to obtain differential dy it is sufficient to multiply it by the given number Δx which is given quite independent of x so that, evidently, no further analytical calculations are necessary. All differentiation laws (both general and special) which we have established in §29 can be converted into laws for evaluation of differentials on simply multiplying the corresponding equation by $\Delta x = dy$. Thus, for example, if $y = \sin x$, we find from

$$y' = \frac{dy}{dx} = \cos x$$

that

$$dy = \cos x \, dx;$$

if $y = \ln x$, we obtain similarly from $y' = dy/dx = 1/x$:

$$dy = \frac{dx}{x},$$

etc. If $y = y_1 + y_2 \pm \dots \pm y_n$, then it follows from the established law that

$$y' = y'_1 \pm y'_2 \pm \dots \pm y'_n,$$

or, which is the same,

$$\frac{dy}{dx} = \frac{dy_1}{dx} \pm \frac{dy_2}{dx} \pm \dots \pm \frac{dy_n}{dx},$$

and on multiplying by dx we obtain :

$$dy = dy_1 \pm dy_2 \pm \dots \pm dy_n$$

(the law for finding differentials of an algebraic sum). On the basis of the corresponding laws for derivatives we can similarly establish differentiation laws for products or functions :

$$d(uv) = u dv + v du,$$

$$d(u_1 u_2 \dots u_n) = du_1 (u_2 \dots u_n) + u_1 du_2 (u_3 \dots u_n) + \\ + u_1 u_2 du_3 (u_4 \dots u_n) + \dots + u_1 u_2 \dots u_{n-1} du_n,$$

$$d\left(\frac{u}{v}\right) = \frac{v du - u dv}{v^2}.$$

For further useful examples cf. problem book by B. P. Demidovich, section II, problem Nos. 151-156.

§ 33. Invariant character of the relationship between a derivative and a differential

We have seen that the differential and the increment are the same for an independent variable and, therefore, if x is the independent variable, the initial expression

$$dy = f'(x) \Delta x \quad (1)$$

for the differential of the function $y = f(x)$ can be written in the form

$$dy = f'(x) dx. \quad (2)$$

Let us now assume that x is not the independent variable but is in its turn an arbitrary (differentiable) function of a new independent variable t :

$$x = \varphi(t).$$

Since the differential and the increment of the *function* (in contrast to the independent variable) are, in general, no longer equal to one another, therefore, in this case $dx \neq \Delta x$, and both relations (1) and (2) can thus no longer be right; only one relation is at best valid. We will show that, irrespective of the nature of the (differentiable) function $\varphi(t)$, the relation (2) always remains valid.

In fact, if $y = f(x)$ and $x = \varphi(t)$, where t is the independent variable, then we can regard $y = f[\varphi(t)]$ as a composite function of t .

As we know, the derivative of this function is equal to $f'(x) \varphi'(t)$ and therefore its differential is equal to

$$dy = f'(x) \varphi'(t) dt; \quad (3)$$

but, on the other hand, we have $x = \varphi(t)$ and therefore

$$dx = \varphi'(t) dt;$$

hence it follows from (3) that

$$dy = f'(x) dx,$$

which was to be proved.

Thus the relation (2), or its equivalent relation

$$y' = f'(x) = \frac{dy}{dx}$$

applies in both cases irrespective of the fact whether x is the independent variable or an arbitrary function of another quantity. This relationship between a derivative and a differential is said to be *invariant* (unalterable) *with respect to any transformation of the independent variable*.

It is interesting to note that in the light of this invariance the law of differentiation of a composite function

$$\frac{dy}{dt} = f'(x) \varphi'(t)$$

can be written in the form

$$\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt}, \quad (4)$$

for we have shown that $f'(x) = dy/dx$; the law obviously appears trivial in this form; however, it would be wrong to prove this law on the basis of the relation (4), for the relation (4) is itself obtained as a corollary of the invariance of the relation (2), in whose proof we already used the law for differentiation of composite functions.

CHAPTER VIII

DERIVATIVES AND DIFFERENTIALS OF HIGHER ORDERS

§ 34. Derivatives of higher orders

The derivative $y' = f'(x)$ of the function $y = f(x)$ is a function of the variable x on which y depends; the problem of differentiating y' can therefore arise. If $y' = f'(x)$ has a derivative, then this derivative is denoted as follows: $y'' = f''(x)$ and known as the *second derivative* or the *derivative of second order* of the function $y = f(x)$. Similarly the derivative of the function y'' , if it exists, is known as the third derivative of the initial function $y = f(x)$; in general, if the n th derivative $y^{(n)} = f^{(n)}(x)$ of the function y exists and is defined and if the function $y^{(n)}$ is differentiable, then its derivative is denoted by $y^{(n+1)} = f^{(n+1)}(x)$ and is known as the $(n+1)$ -th derivative (or as the derivative of $(n+1)$ -th order) of the initial function $y = f(x)$.

Derivatives of higher orders occur quite frequently in many problems of accurate natural study, technical processes and other scientific and practical branches. Therefore the ability to find them and the knowledge of their properties is not only necessary for mathematicians but also for scientists in every branch where mathematical analysis finds application. We have seen in § 26 that if $s = f(t)$ is the law of motion of a body, then $s' = f'(t)$ expresses the instantaneous velocity $v(t)$ of this motion at the instant t . The second derivative $s'' = f''(t) = v'(t)$, i.e. the derivative of the instantaneous velocity, denotes the "rate of change of velocity"; in mechanics this quantity is known as *acceleration*; its importance is very great mainly because, according to the well-known Newton's law, acceleration is proportional to the acting force; the majority of mechanical problems are phrased in such a way that the acting forces are given and the motion due to their action is to be found; but giving acceleration is equivalent to giving of the acting force and therefore a typical mechanical problem

involves establishment of the character of motion from the given acceleration. The second derivative has several important geometrical applications which we shall learn later.

It is obvious that in order to find derivatives of higher orders it is only necessary to perform successively a series of ordinary differentiations and so no other new methods are needed. Here we shall only note several interesting results for some elementary functions.

1. We have seen in § 29 that derivative of a polynomial $y = a_0x^n + a_1x^{n-1} + \dots + a_n$ is a polynomial of degree one less than the degree of the given polynomial and has the leading term na_0x^{n-1} ; each new differentiation lowers the degree by one each time; thus the derivative of the n th order

$$y^{(n)} = n! a_0$$

is a polynomial of order 0, i.e. a constant; therefore

$$y^{(n+1)} = y^{(n+2)} = \dots = 0,$$

i.e. for a polynomial of the n th degree the derivatives of all orders greater than n are identically equal to zero.

2. We know that the function $y = e^x$ remains unchanged by differentiation ($y' = y$). Therefore evidently $y^{(n)} = y = e^x$ for any n . More generally, if $y = a^x$, we have $y' = y \ln a$, and therefore $y^{(n)} = y (\ln a)^n = a^x (\ln a)^n$ for any n .

3. The derivative of the function $y = \sin x$ is $y' = \cos x$ and that of the function $z = \cos x$ is $z' = -\sin x$; hence

$$y'' = -\sin x, y''' = -\cos x, y^{(4)} = \sin x, y^{(5)} = \cos x, \dots;$$

the successive derivatives of the function $\sin x$ form, as we can see, a periodic series with a period of 4, so that for any n

$$y^{(4n)} = \sin x, y^{(4n+1)} = \cos x, y^{(4n+2)} = -\sin x, y^{(4n+3)} = -\cos x;$$

and similarly for the function $z = \cos x$:

$$z^{(4n)} = \cos x, z^{(4n+1)} = -\sin x, z^{(4n+2)} = -\cos x, z^{(4n+3)} = \sin x;$$

this is the same series as above but it has shifted by one position.

4. The derivatives of the functions $\ln x$, $\arctan x$ and $\operatorname{arccot} x$ are, as we know, expressed in terms of rational fractions; therefore it evidently follows that the derivatives of all orders of these functions will also be rational fractions; similarly the derivatives of all orders of the functions $\arcsin x$ and $\arccos x$ are algebraic functions.

5. We know that the first derivative of every elementary function is in general also an elementary function; it therefore evidently follows that the derivatives of all orders of elementary functions are always elementary functions.

6. The law of differentiation of an algebraic sum can evidently be extended without changes to derivatives of all orders. However, the second differentiation of product of two functions deserves special attention: if $y = uv$, where u and v are differentiable functions of x , then, as we know,

$$y' = uv' + vu',$$

and it can readily be seen that

$$y'' = uv'' + 2u'v' + u''v,$$

$$y''' = uv''' + 3u'v'' + 3u''v' + u'''v,$$

which makes it reasonable to assume that for every n

$$y^{(n)} = a_{n0}uv^{(n)} + a_{n1}u'v^{(n-1)} + a_{n2}u''v^{(n-2)} + \dots \\ \dots + a_{n,n-1}u^{(n-1)}v' + a_{nn}u^{(n)}v, \quad (1)$$

where $\alpha_{n0}, \alpha_{n1}, \dots, \alpha_{nn}$ are constants independent of the form of the functions u and v ; this proposition which we have already proved for $n=1, 2$ and 3 , can readily be proved, for every n by the method of induction, (we leave the proof to the reader). The numbers $\alpha_{n0}, \alpha_{n1}, \dots, \alpha_{nn}$ must only be found. Since these numbers are independent of the form of the functions u and v , they can be chosen specially. Assuming that

$$u = e^x, v = e^{tx},$$

where t is an arbitrary constant, we find that

$$u^{(n)} = e^x, v^{(n)} = t^n e^{tx},$$

$$y = e^{(t+1)x}, y^{(n)} = (t+1)^n e^{(t+1)x},$$

and formula (1) gives :

$$(t+1)^n e^{(t+1)x} = \alpha_{n0} e^x t^n e^{tx} + \alpha_{n1} e^x t^{n-1} e^{tx} + \\ + \alpha_{n2} e^x t^{n-2} e^{tx} + \dots + \alpha_{nn} e^x e^{tx} = \\ = e^{(t+1)x} (\alpha_{n0} t^n + \alpha_{n1} t^{n-1} + \alpha_{n2} t^{n-2} + \dots + \alpha_{nn}),$$

and therefore

$$(t+1)^n = \alpha_{n0} t^n + \alpha_{n1} t^{n-1} + \alpha_{n2} t^{n-2} + \dots + \alpha_{nn},$$

where the number t is arbitrary ; comparing this formula with the expansion of $(t + 1)^n$ by the binomial formula we can see that two polynomials which are identically equal should have similar corresponding coefficients in pairs and therefore

$$\alpha_{nk} = C_n^k \quad (k = 0, 1, \dots, n)$$

and formula (1) gives :

$$y^{(n)} = C_n^0 u v^{(n)} + C_n^1 u' v^{(n-1)} + \dots + C_n^{n-1} u^{(n-1)} v' + C_n^n u^{(n)} v.$$

This is the so-called *Leibnitz formula* which gives the n th derivative of a product of two functions in terms of the derivatives of the factors up to n th order inclusively.

§ 35. Differentials of higher orders and their relationship with derivatives

Differentials of higher orders are determined just as derivatives. The second differential d^2y of the function $y = f(x)$ is the differential of the first differential

$$d^2y = d(dy);$$

and, in general, if the differential $d^n y$ of order n of the function y is already determined, then

$$d^{n+1}y = d(d^n y).$$

Since dy is by definition a function of two independent variables, viz. x and Δx , the expression $d(dy)$, with whose help the second differential d^2y is determined, requires some explanation ; while performing the operation $d(dy)$ we are always considering dy as a function of x alone by assuming Δx to be constant ; this remark also applies to all subsequent differentials and Δx is assumed to be one and the same for differentials of all orders.

In order to establish a relationship between differentials of higher orders and the corresponding derivatives let us at first recall that

$$dy = y' \Delta x,$$

i.e. the formation of a differential of a given function y involves multiplication of its derivative with respect to x and the increment Δx of the independent variable and, as we have already emphasized

on many occasions, the quantities x and Δx should be regarded as independent of one another. Thus in order to find the second differential $d^2 y = d(dy)$ of the function y we must find the derivative dy with respect to x and multiply it by Δx . But $dy = y' \Delta x$, where the second factor is independent of x and in differentiation of product with respect to x it should be regarded as a constant; the derivative of $dy = y' \Delta x$ with respect to x is therefore equal to $y'' \Delta x$ and hence

$$d^2 y = d(dy) = y'' (\Delta x)^2;$$

repetition of this operation evidently gives

$$d^3 y = y''' (\Delta x)^3,$$

and in general

$$d^n y = y^{(n)} (\Delta x)^n;$$

the differential of order n is equal to the derivative of the same order multiplied by the n th power of the increment Δx . Conversely, it gives

$$y^{(n)} = \frac{d^n y}{(\Delta x)^n},$$

or, if we remember that $\Delta x = dx$,

$$y^{(n)} = \frac{d^n y}{dx^n} \quad (1)$$

where the denominator should be regarded as $(dx)^n$ but, for the sake of simplicity of notation, brackets are always omitted. Thus *the derivative of order n is equal to the differential of the same order divided by the n th power of the (first) differential of the independent variable.*

Formula (1) is a generalisation of the formula $y' = dy/dx$ and, like this formula, can in many instances serve as a very convenient method of notation for derivatives of higher orders. However, the formula $y' = dy/dx$ is, as we know, invariant with respect to every transformation of the independent variable (*i.e.* it remains valid even when x is not the independent variable but a function of a new variable t); formula (1), no longer possesses this invariance property for $n > 1$, and essentially depends on the fact that x is the independent variable. In fact, we will show that formula (1) is in general not valid for $n = 2$ if $x = \varphi(t)$. We know that in this case (assuming that $y = f(x)$)

$$dy = f'(x) dx = f'[\varphi(t)] \varphi'(t) dt.$$

While taking the second differential $d^2y = d(dy)$ we should differentiate dy with respect to t and multiply the result by dt . This gives

$$\begin{aligned} d^2y &= \{f''[\varphi(t)]\varphi'(t)^2 + f'[\varphi(t)]\varphi''(t)\} dt^2 = \\ &= f''[\varphi(t)] [\varphi'(t)dt]^2 + f'[\varphi(t)]\varphi''(t)dt^2 = f''(x)dx^2 + f'(x)d^2x, \end{aligned}$$

since $\varphi'(t)dt = dx$ and $\varphi''(t)dt^2 = d^2x$. We therefore obtain

$$\frac{d^2y}{dx^2} = f''(x) + f'(x) \frac{d^2x}{dx^2},$$

whereas if x is the independent variable we have

$$\frac{d^2y}{dx^2} = f''(x);$$

the additional term

$$f'(x) \frac{d^2x}{dx^2}$$

appears as a result of the fact that x is now a function of the independent variable t ; in fact, if x is the independent variable, then $dx = \Delta x$, $d^2x = 0$ and the additional term is absent.

We have already said that finding differentials and derivatives of higher orders does not necessitate any fundamentally new methods and therefore not many exercises are needed. The student will find many interesting problems in the problem book by B. P. Demidevich, section II, §5.

CHAPTER IX.

MEAN VALUE THEOREMS

§ 36. Theorem on finite increments

In the last three chapters we have studied practical operations of differentiation. We have learnt how to find derivatives and differentials, and the theorems proved there are mainly designed to help and facilitate this process. As we have done all this and learnt the technique of differentiation, we must study some further properties of derivatives and differentials, *i.e.* we must study the properties which form the theoretical basis of differential calculus. Among the laws which we shall study as such a prominent part is played by several theorems which can be given the common name “mean value theorems”; this title in general involves propositions which imply existence, under specific conditions, of a given point c (or “mean value”) in the given interval (a, b) at which the given function possesses certain properties. We have already met one such theorem in Chapter 5 (theorem 3 §23) : if the function $f(x)$ is continuous in the interval (a, b) and has opposite signs on its ends, then a point c can be found in this section such that $f(c) = 0$. The main characteristic of theorems of this type is due to the fact that they do not give any indications as to the position of the point c in the interval (a, b) but only prove the mere fact of its existence. We shall now establish several such theorems for the function $f(x)$ which is differentiable at every point in the interval (a, b) and in every case we shall assume that $\lim \Delta y / \Delta x$ exists at the point a only for $\Delta x \rightarrow +0$ and at the point b only for $\Delta x \rightarrow -0$,

Let us at first prove an auxiliary proposition which we shall find useful later.

Lemma. *If the function $f(x)$ has a derivative at the point x and if the following inequalities holds*

$$f(x + h) \leq f(x), \quad f(x - h) \leq f(x) \quad (1)$$

for all sufficiently small $h > 0$, then $f'(x) = 0$.

Proof. It is given that $f'(x)$ exists; therefore for $h \rightarrow +0$ we should have :

$$\frac{f(x + h) - f(x)}{h} \rightarrow f'(x), \quad \frac{f(x - h) - f(x)}{-h} \rightarrow f'(x);$$

the first of these fractions, as a result of the statement of the lemma, cannot be positive for a sufficiently small h and therefore (corollary 2, theorem 2 §10) its limit must also be $f'(x) \leq 0$; similarly the second fraction cannot be negative for a sufficiently small h and therefore its limit must be $f'(x) \geq 0$; hence the derivative $f'(x)$ can neither be positive nor negative and it must therefore be equal to zero.

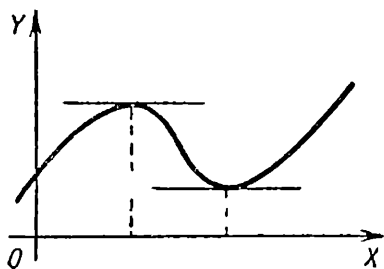


Fig. 16.

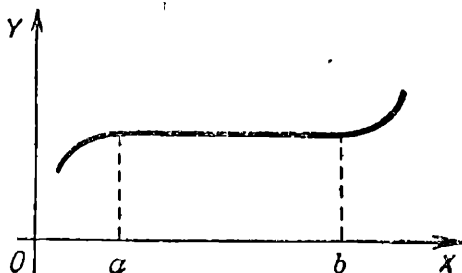


Fig. 17.

This lemma means that *at the point where the function acquires its maximum value as compared to an adjacent value, the derivative, in case it exists, should be equal to zero.* The lemma evidently remains valid also when the value of the function $f(x)$ attains its *minimum* at the point x as compared to an adjacent value, *i.e.* when the inequalities (1) are replaced by opposite inequalities. This lemma can be geometrically illustrated by representing $y = f(x)$ in the graphical form (Fig. 16) : at the point where the curve $y = f(x)$ attains its highest or lowest position as compared with its immediate neighbourhood, the tangent, in case it exists, should be parallel to the OX -axis; in this event we do not exclude the case when the function also has the same values (as at the point x) at other points which can be as close to x as we please (or which can be sufficiently close to x); thus for the function represented in Fig. 17 the statement expressed by the lemma remains valid for every point on the line (a, b) .

Theorem (Rolle's). *If $f(a) = f(b)$ and the function $f(x)$ is continuous in the interval (a, b) and differentiable at every point in that interval, then an interior point c can be found in the interval (a, b) at which $f'(c) = 0$.*

Proof. Let us assume that $f(a) = f(b) = \gamma$ for all points x in the interval (a, b) , i.e., the function $f(x)$ is constant in that interval, then $f'(x) = 0$ at every point x in that interval. Otherwise the interval (a, b) will contain points at which $f(x) > \gamma$ or other points at which $f(x) < \gamma$ (it may, of course, happen that neither points exist). For the sake of argument let us assume that there exist points for which $f(x) > \gamma$.

Since the function $f(x)$ is continuous in the interval (a, b) , therefore, according to theorem 2 § 23, it should attain its maximum

value at a point c in that interval; it is evident that $f(c) > \gamma$; therefore the point c does not coincide with a or b , i.e. it is an interior point of the interval (a, b) ; it follows from the definition of the point c that for all points in the interval (a, b) , including all the points x situated sufficiently close to the point c , we have $f(x) \leq f(c)$; hence applying the lemma we have $f'(c) = 0$ and the theorem is thus proved.

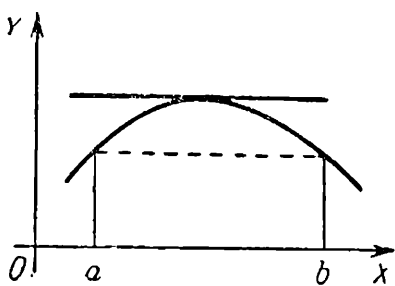


Fig. 18.

The geometrical illustration of Rolle's theorem is evidently based on the fact that between two points on the given curve, situated on the same level, a point can always be found so that the tangent at this point is horizontal (Fig. 18); in this case we assume that a tangent at every point exists on the given section of the curve.

Theorem (Lagrange, on finite increments). *If the function $f(x)$ is continuous in the interval (a, b) and differentiable at every interior point of this interval, then a point c can be found in this interval at which*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \quad (2)$$

Owing to the fact that

$$\frac{f(b) - f(a)}{b - a}$$

is the slope of the chord joining the points $[a, f(a)]$ and $[b, f(b)]$ of the

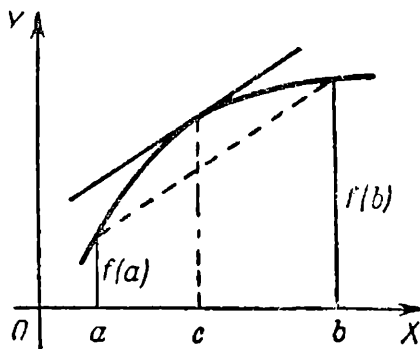


Fig. 19.

curve $y = f(x)$ (Fig. 19), therefore from a geometrical point of view Lagrange's theorem maintains that for a curve which has a tangent at every point, a point can be found between the ends of every chord at which the tangent will be parallel to the chord. It is evident that Rolle's theorem is a particular case of Lagrange's theorem when the given chord is parallel to the OX -axis.

It is obvious from the geometrical representation that the general case can be deduced from the particular one simply by turning the diagram and therefore the analytical proof should also not be complicated if we base it on Rolle's theorem.

Proof. Let us consider the auxiliary function

$$\varphi(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a} (x - a),$$

which geometrically represents the difference between the ordinate of a curve and ordinate of the chord as shown in Fig. 19. Evidently $\varphi(a) = \varphi(b) = 0$; on the other hand the function $\varphi(x)$, like the function $f(x)$, is continuous in interval (a, b) and differentiable at every point of this interval, and

$$\varphi'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

It follows from Rolle's theorem that a point c can be found in the interval (a, b) such that

$$\varphi'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0,$$

which prove Lagrange's theorem.

This is one of the most important theorems of differential calculus and we shall frequently use it. The relation (2) stated by this theorem is sometimes, for the sake of convenience, written in the form

$$f(b) - f(a) = f'(c) (b - a). \quad (3)$$

The meaning of this statement is not altered in any way by the fact that if the function $f(x)$ has a derivative at every point in the interval (a, b) , then a point c can be found between a and b at which the relation (3) holds.

Let us finally rewrite the same relation in a different notation. Let us write x instead of a and $x + \Delta x$ instead of b , where $b - a = \Delta x$; we thus obtain :

$$f(x + \Delta x) - f(x) = f'(c) \Delta x \quad (x < c < x + \Delta x).$$

If we denote $f(x)$ by y , then it is convenient to denote the left-hand side of this relation by Δy , as we did in the past. It is also convenient to denote the point c , about which we only know that it lies between x and $x + \Delta x$, by $x + \theta \Delta x$, if we agree that θ denotes a number (unknown) which lies between 0 and 1 ($0 < \theta < 1$). Our relation thus becomes

$$\Delta y = f(x + \Delta x) - f(x) = f'(x + \theta \Delta x) \Delta x. \quad (4)$$

It is interesting to compare this equation with another equation which we have used several times in chapter 7 :

$$\Delta y = f'(x) \Delta x + o(\Delta x);$$

this relation shows that the increment Δy of the function $y = f(x)$ is equal to product $f'(x) \Delta x$ with an accuracy up to infinitely small quantities of higher orders; the equation (4) (i.e. the theorem on finite increments) shows that in this expression the term $o(\Delta x)$ can be completely rejected but that in the main term the derivative $f'(x)$ must be replaced by a derivative at some (unknown) point $x + \theta \Delta x$ which lies between x and $x + \Delta x$; both relations are very useful and have many applications.

The following theorem is an important generalisation of Lagrange's theorem :

Theorem (Cauchy). *If the functions $f(x)$ and $\varphi(x)$ are continuous in the interval (a, b) and differentiable at every point in that interval where $\varphi'(x) \neq 0$ ($a < x < b$), then a point c ($a < c < b$) exists such that*

$$\frac{f(b) - f(a)}{\varphi(b) - \varphi(a)} = \frac{f'(c)}{\varphi'(c)}, \quad (5)$$

(i.e. the ratio of increments of two functions is equal to the ratio of their derivatives at one and the same point in a given interval).

The proof of this theorem can be carried out in exactly the same way as the proof of Lagrange's theorem. The following function should be taken as the auxiliary function

$$f(x) - f(a) - \frac{f(b) - f(a)}{\varphi(b) - \varphi(a)} [\varphi(x) - \varphi(a)]$$

$[\varphi(b) - \varphi(a) \neq 0$, since otherwise, in accordance with Rolle's theorem, we would have $\varphi'(c) = 0$ at a point c ($a < c < b$) which contradicts the conditions of the theorem]; all other arguments are the same as above and give us the relation

$$f'(c) - \frac{f(b) - f(a)}{\varphi(b) - \varphi(a)} \varphi'(c) = 0 \quad (a < c < b),$$

from which (5) follows.

It is evident that Cauchy's theorem becomes Lagrange's theorem when we choose $\varphi(x) = x$ and is, in fact, a generalisation of this theorem.

We have already said that the theorems studied in this paragraph have many applications in analysis. We shall now consider a simple but very important example of this kind of application.

We know that the derivative of a constant is equal to zero. Is the converse also true, *i.e.* can we say that the function $f(x)$ whose derivative at every point of a given interval is equal to zero is constant in that interval? To answer this question let us take two arbitrary points x_1 and x_2 in the given interval; it follows from Lagrange's theorem that

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1),$$

where c is a point between x_1 and x_2 ; but we have assumed that $f'(x) = 0$ at every point x in our interval and therefore also $f'(c) = 0$, and hence $f(x_2) = f(x_1)$; the values of the function $f(x)$ are equal at two arbitrary points in the given interval and this means that the function $f(x)$ is constant in that interval. We thus see that the theorem on finite increments enables us to prove the following important proposition which we shall frequently use:

Theorem. *If $f'(x) = 0$ at every point in the interval (a, b) , then the function $f(x)$ is constant in that interval.*

In the next two paragraphs we shall consider some other important applications of the proved mean value theorems.

§ 37. Evaluation of limits of ratios of infinitely small and infinitely large quantities

While considering the general theory of limits (chapter 2) we have said that the ratio of two infinitely small quantities (of infinitely large quantities) can in a given process have a very diverse character of change in relation to the nature of the infinitely small (or infinitely large) quan-

tities so that we cannot make any general predictions as to the behaviour of ratios of this type. At the same time the true value of these ratios is very important : thus, as we now know, the derivative of a given function, which is the fundamental concept in differential calculus, is defined as limit of a ratio of infinitesimals. It is therefore clear how valuable is to find a more or less general method for evaluating limits of such ratios in case they exist. One such very useful and at the same time simple and powerful method can be developed on the basis of mean value theorems which have been proved in the previous paragraph. We shall now consider this method.

Let us assume that the point a belongs (i.e. it lies on or is one of the ends) to a segment Δ along which the functions $f_1(x)$ and $f_2(x)$ are continuous ; let $f_1(a) = f_2(a) = 0$, where both functions are differentiable at every point in this segment and $f'_2(x) \neq 0$. In that case $f_2(x) \neq 0$ ($x \neq a$) along Δ , for otherwise, in accordance with Rolle's theorem, $f'_2(x)$ would vanish at a point in the segment Δ other than the point a . We can therefore consider the ratio $f_1(x)/f_2(x)$ of two infinitely small quantities and try to find its limit for $x \rightarrow a$. Since $f_1(a) = f_2(a) = 0$, therefore

$$\frac{f_1(x)}{f_2(x)} = \frac{f_1(x) - f_1(a)}{f_2(x) - f_2(a)} ;$$

it follows from Cauchy's theorem proved in § 26 that all requirements are evidently satisfied in this case and therefore

$$\frac{f_1(x)}{f_2(x)} = \frac{f'_1(c)}{f'_2(c)} , \quad (1)$$

where c is some point ("mean value") between a and x . Let it now be known that the ratio $f'_1(x) / f'_2(x)$ tends to a certain limit l for $x \rightarrow a$; since c lies between a and x , therefore for $x \rightarrow a$ and $c \rightarrow a$ we have

$$\frac{f'_1(c)}{f'_2(c)} \rightarrow l \quad (x \rightarrow a) ,$$

and equation (1) also shows that

$$\frac{f_1(x)}{f_2(x)} \rightarrow l \quad (x \rightarrow a).$$

We have thus proved a theorem known as l'Hospital's rule :

Let $f_1(a) = f_2(a) = 0$ and let the functions $f_1(x)$ and $f_2(x)$ be continuous along a segment Δ on which the point a is situated; if in this case

$f'_1(x)$ and $f'_2(x)$ exist for all points $x \neq a$ along the segment Δ , $f'_2(x) \neq 0$ ($x \neq a$) and $f'_1(x)/f'_2(x) \rightarrow l$ for $x \rightarrow a$, then $f_1(x)/f_2(x) \rightarrow l$ for $x \rightarrow a$.

The significance of this rule is due to the fact that in many cases the limit of the ratio of derivatives can be found much more easily than the limit of the ratio of the functions themselves; it may happen in some cases that one or other derivative is no longer infinitely small for $x \rightarrow a$; in that case we are no longer dealing a ratio of infinitely small quantities and the limit can be found quite easily.

Example 1. When $b \neq 0$, $x \rightarrow 0$,

$$\lim_{x \rightarrow 0} \frac{\sin ax}{\sin bx} = \lim_{x \rightarrow 0} \frac{a \cos ax}{b \cos bx} = \frac{a}{b}.$$

Example 2. When $x \rightarrow 0$, $\lim_{x \rightarrow 0} \frac{\tan x - x}{x - \sin x} = \lim_{x \rightarrow 0} \frac{\frac{1}{\cos^2 x} - 1}{1 - \cos x} =$
 $= \lim_{x \rightarrow 0} \frac{1 - \cos^2 x}{\cos^2 x (1 - \cos x)} = \lim_{x \rightarrow 0} \frac{1 + \cos x}{\cos^2 x} = 2.$

Many other useful exercises can be found in the problem book by B.P. Demidovich, Section II, § 10.

If both derivatives $f'_1(x)$ and $f'_2(x)$ are infinitely small for $x \rightarrow a$ and are in their turn differentiable in the neighbourhood of a point a (where $f''_2(x)$ does not vanish for $x \neq a$), there is no reason why l'Hospital's rule should not be applied again: if we have $f''_1(x)/f''_2(x) \rightarrow l$, for $x \rightarrow a$, then according to this rule we also have $f'_1(x)/f'_2(x) \rightarrow l$ and therefore also $f_1(x)/f_2(x) \rightarrow l$ for $x \rightarrow a$. In general, if the functions $f_1(x)$ and $f_2(x)$ have derivatives of order n in some neighbourhood of the point a , where $f_2^{(n)}(x) \neq 0$ for $x \neq a$ and $f_1(a) = f_2(a) = f'_1(a) = f'_2(a) = \dots = f_1^{(n-1)}(a) = f_2^{(n-1)}(a) = 0$, then by applying l'Hospital's rule for the second time we can evidently conclude that if the following relation exists

$$\lim_{x \rightarrow a} \frac{f_1^{(n)}(x)}{f_2^{(n)}(x)} = l,$$

then the limit of the ratio $f_1(x)/f_2(x)$ for $x \rightarrow a$ also exists and is equal to l *).

*) The relation necessary for the application of l'Hospital's rule for the second time, i.e. $f_2^{(k)}(x) \neq 0$ ($1 \leq k \leq n$, $x \neq a$) for $k = n$ is one of the assumptions of the theorem and can readily be established for lower values of k by the method of induction, applying Rolle's theorem to the function $f_2^{(k)}(x)$ along the line (a, x) .

Example 3. $f_1(x) = x - \sin x$, $f_2(x) = x^3$; we have

$$f'_1(x) = 1 - \cos x, f''_1(x) = \sin x, f'''_1(x) = \cos x,$$

$$f'_2(x) = 3x^2, f''_2(x) = 6x, f'''_2(x) = 6,$$

and therefore

$$f_1(0) = f'_1(0) = f''_1(0) = 0, \quad f'''_1(0) = 1,$$

$$f_2(0) = f'_2(0) = f''_2(0) = 0, \quad f'''_2(0) = 6,$$

$$\frac{f'''_1(x)}{f'''_2(x)} = \frac{x - \sin x}{x^3} \rightarrow \frac{1}{6} \quad (x \rightarrow 0)$$

and hence also

$$\frac{f_1(x)}{f_2(x)} = \frac{x - \sin x}{x^3} \rightarrow \frac{1}{6} \quad (x \rightarrow 0).$$

L'Hospital's rule remains valid when $x \rightarrow \infty$. Thus, for example, if the functions $f_1(x)$ and $f_2(x)$ are infinitely small for $x \rightarrow +\infty$ and differentiable for sufficiently large x and $f'_2(x) \neq 0$, then it follows from

$$\frac{f'_1(x)}{f'_2(x)} \rightarrow l \quad (x \rightarrow +\infty) \quad (2),$$

that we also have

$$\frac{f_1(x)}{f_2(x)} \rightarrow l \quad (x \rightarrow +\infty)$$

In fact, assuming that $x = 1/y$ we have:

$$\frac{f_1(x)}{f_2(x)} = \frac{f_1\left(\frac{1}{y}\right)}{f_2\left(\frac{1}{y}\right)} = \frac{\varphi_1(y)}{\varphi_2(y)},$$

and for $y \rightarrow +0$

$$\varphi_1(y) \rightarrow 0, \quad \varphi_2(y) \rightarrow 0.$$

Since

$$\frac{\varphi'_1(y)}{\varphi'_2(y)} = \frac{f'_1\left(\frac{1}{y}\right)\left(-\frac{1}{y^2}\right)}{f'_2\left(\frac{1}{y}\right)\left(-\frac{1}{y^2}\right)} = \frac{f'_1\left(\frac{1}{y}\right)}{f'_2\left(\frac{1}{y}\right)},$$

it follows from (2) that

$$\frac{\varphi'_1(y)}{\varphi'_2(y)} \rightarrow l \quad (y \rightarrow +0);$$

and therefore on the basis of l'Hospital's rule we have

$$\frac{\varphi_1(y)}{\varphi_2(y)} \rightarrow l \quad (y \rightarrow +0);$$

and this is equivalent to the relation

$$\frac{f_1(x)}{f_2(x)} \rightarrow l \quad (x \rightarrow +\infty).$$

Example 4. When $x \rightarrow +\infty$,

$$\begin{aligned} \lim x \left(\frac{\pi}{2} - \arctan x \right) &= \lim \frac{\frac{\pi}{2} - \arctan x}{\frac{1}{x}} = \lim \frac{-\frac{1}{1+x^2}}{-\frac{1}{x^2}} = \\ &= \lim \frac{x^2}{x^2 + 1} = 1. \end{aligned}$$

Let us now consider a ratio of two infinitely large quantities. We shall see that l'Hospital's rule also remains valid in this case, although the proof is somewhat more complicated. Let us assume that we have

$$|f_1(x)| \rightarrow +\infty, \quad |f_2(x)| \rightarrow +\infty \quad (x \rightarrow a),$$

and let, as before, both functions be differentiable in a neighbourhood of the point a , where $f'_2(x) \neq 0$ for all $x \neq a$. Let us take two points x and α in this neighbourhood, both situated on the same side of the point a so that, for example, $a < x < \alpha$. It follows from Cauchy's theorem that

$$\frac{f_1(x) - f_1(\alpha)}{f_2(x) - f_2(\alpha)} = \frac{f'_1(c)}{f'_2(c)},$$

where $x < c < \alpha$. But on the other hand

$$\frac{f_1(x) - f_1(\alpha)}{f_2(x) - f_2(\alpha)} = \frac{f_1(x)}{f_2(x)} \frac{1 - \frac{f_1(\alpha)}{f_1(x)}}{1 - \frac{f_2(\alpha)}{f_2(x)}};$$

and comparison of these two equations gives:

$$\frac{f_1(x)}{f_2(x)} = \frac{f'_1(c)}{f'_2(c)} \frac{1 - \frac{f_2(\alpha)}{f_2(x)}}{1 - \frac{f_1(\alpha)}{f_1(x)}}. \quad (3)$$

Let us now assume that the following limit exists :

$$\lim_{x \rightarrow a} \frac{f'_1(x)}{f'_2(x)} = l.$$

Let $\varepsilon > 0$ be as small as we please; let us choose α so close to a that for $a < z < \alpha$

$$\left| \frac{f'_1(z)}{f'_2(z)} - l \right| < \varepsilon,$$

or, which is the same,

$$l - \varepsilon < \frac{f'_1(z)}{f'_2(z)} < l + \varepsilon,$$

so that the point c should lie between a and α and

$$l - \varepsilon < \frac{f'_1(c)}{f'_2(c)} < l + \varepsilon \quad (4)$$

(as x changes, c will also change, but since it always remains confined between a and α , the inequalities (4) remain valid). Let us now assume that while α remains unchanged, x tends to a ; it then follows from our assumptions that in this case $|f_1(x)| \rightarrow +\infty$ and $|f_2(x)| \rightarrow +\infty$, and therefore the second factor on the right-hand side of the relation (3) will tend to unity; it can be represented in the form $1 + \delta$, where $\delta \rightarrow 0$ for $x \rightarrow a$. Multiplying both sides of the inequalities (4) by this factor we obtain as a result of the relation (3):

$$(1 + \delta)(l - \varepsilon) < \frac{f_1(x)}{f_2(x)} < (1 + \delta)(l + \varepsilon);$$

since the number ε is as small as we please and $\delta \rightarrow 0$ for $x \rightarrow a$, it evidently follows that

$$\frac{f_1(x)}{f_2(x)} \rightarrow l \quad (x \rightarrow a),$$

which was to be proved.

Example 5. $\ln x \rightarrow -\infty$ for $x \rightarrow 0$. Hence it cannot be immediately seen as to how $x \ln x$ behaves. In order to study this behaviour we note that

$$-x \ln x = \frac{-\ln x}{\frac{1}{x}}$$

can be represented as a ratio of two infinitely large quantities; the derivatives of the numerator and denominator are respectively equal to $-1/x$ and $-1/x^2$ and their ratio is equal to x and tends to zero; it therefore follows from l'Hospital's rule that

$$x \ln x \rightarrow 0 \quad (x \rightarrow 0).$$

As before, it could readily be proved that l'Hospital's rule remains valid for the ratio of two infinitely large quantities when x does not tend to a finite limit but increases indefinitely.

Example 6. For $x \rightarrow +\infty$

$$\lim \frac{\ln x}{\sqrt{x}} = \lim \frac{\frac{1}{x}}{\frac{1}{2\sqrt{x}}} = \lim \frac{2}{\sqrt{x}} = 0$$

and generally ($\alpha > 0$) for $x \rightarrow \infty$

$$\lim \frac{\ln x}{x^\alpha} = \lim \frac{\frac{1}{x}}{\alpha x^{\alpha-1}} = \lim \frac{1}{\alpha x^\alpha} = 0.$$

Example 7. Let $a > 1$, $\alpha > 0$. The functions x^α and a^x increase indefinitely for $x \rightarrow +\infty$. Let n be the greatest integer smaller than α , so that $0 \leq n < \alpha \leq n+1$. It can readily be seen that all derivative of the function x^α , up to and including the n th order increase indefinitely for $x \rightarrow \infty$, whereas the derivative of order $n+1$ is equal to $\alpha(\alpha-1) \dots (\alpha-n) x^{\alpha-n-1}$ and remains bounded. Since the $n+1$ th derivative of the function a^x is equal to $a^x (\ln a)^{n+1}$ and increases indefinitely for $x \rightarrow +\infty$, therefore the application of l'Hospital's rule $n+1$ times shows that

$$\frac{x^\alpha}{a^x} \rightarrow 0 \quad (x \rightarrow +\infty)$$

for every $\alpha > 0$ and $a > 1$.

§ 38. Taylor's Formula

We shall base our argument on the well-known relation established in § 31 : if the function $f(x)$ has a derivative at the point a , then

$$f(a + h) = f(a) + f'(a) h + o(h) \quad (1)$$

for $h \rightarrow 0$ when $|h|$ is small; this relation enables us to express approximately the value of $f(a + h)$ in terms of a linear function, although this function has usually a rather complicated dependence on h

$$f(a + h) \approx f(a) + f'(a) h,$$

where the error of this approximate evaluation is of the form $o(h)$, *i.e.* this error is negligibly small for small values of h , not only in itself but also as compared to $|h|$. We have already learnt that this fact has a very great practical value, since it makes possible to find very readily a good approximation for the value of $f(a + h)$ (*c.f.* § 31). We shall now learn that this point serves as a basis for further development of the theory.

About the quantity $o(h)$ in equation (1) we know only that it is an infinitely small quantity of a higher order as compared to h ; we have no other accurate information about it. Hence the question how far formula (1) is suitable for approximate evaluation of $f(a + h)$ depends entirely on the desired degree of accuracy. If the required degree of accuracy is such that a quantity of the type $o(h)$ (*i.e.* an infinitely small quantity of a higher order as compared to h) can be neglected, then formula (1) solves our problem; otherwise it is not sufficiently accurate. It may happen (and it does frequently so) that we are obliged to take into account infinitely small quantities of second order with respect to h [*i.e.* quantities of the same order as h^2]; but we can disregard all quantities above the second order (*i.e.* quantities of the type $o(h^2)$). In that case we shall look for a more accurate expression for $f(a + h)$ and use a formula similar to formula (1)

$$f(a + h) = \alpha_0 + \alpha_1 h + \alpha_2 h^2 + o(h^2),$$

where $\alpha_0, \alpha_1, \alpha_2$ are constants (independent of h), *i.e.* we shall look for the approximate value of $f(a + h)$ in the form of a trinomial of second degree

$$f(a + h) \approx \alpha_0 + \alpha_1 h + \alpha_2 h^2,$$

in which the error is of the type $o(h^2)$, *i.e.* an infinitely small quantity above the second order as compared to h . At first we evidently know nothing about the existence of such a polynomial and we have no way of finding its coefficients $\alpha_0, \alpha_1, \alpha_2$; therefore all that we have said in this connection can merely be regarded as the statement of the problem.

However, before attempting to solve this problem we will naturally state it in a more general form. The real nature of the problem for which we are trying to find an approximate value of the function $f(a + h)$ determines the required degree of accuracy. By making an assumption of a fairly general character we can then decree that quantities of the order h^n (where n is a constant natural number) should still be taken into account, but infinitely small quantities of higher orders (*i.e.* quantities of the type $o(h^n)$) should be neglected. The question arises whether 1) a polynomial of the n th degree exists

$$P_n(h) = \alpha_0 + \alpha_1 h + \alpha_2 h^2 + \dots + \alpha_n h^n$$

(with coefficients independent of h) so that

$$f(a + h) - P_n(h) = o(h^n) \quad (2)$$

for $h \rightarrow 0$ and 2) if it does exist, then how we can find its coefficients. If the problem so stated can be satisfactorily solved, then the polynomial $P_n(h)$ will enable us to find the value of $f(a + h)$ with the required degree of accuracy: for practical calculations (and also for theoretical investigations) we know nothing more convenient and simple than a polynomial.

It can, of course, be foreseen that the answer to the above questions will depend to a large extent on the nature of the function $f(x)$ in the neighbourhood of the point a . Already in the case considered earlier for $n = 1$ we had to introduce the condition of differentiability of the function $f(x)$ at the point a . If we want to express approximately the value of $f(a + h)$ in terms of a polynomial of the n th degree with an accuracy up to quantities of the type $o(h^n)$, we shall have to assume that the function $f(x)$ has derivatives of all orders up to the n th order inclusively at the point a [in other words, we must assume existence of $f^{(n)}(a)$]. But this will be the only assumption we shall have to make.

We will now show that if $f^{(n)}(a)$ exists, then we have

$$\begin{aligned} f(a+h) &= f(a) + f'(a)h + \frac{1}{2!}f''(a)h^2 + \dots \\ &\quad \dots + \frac{1}{n!}f^{(n)}(a)h^n + o(h^n) \end{aligned} \quad (T)$$

for $h \rightarrow 0$; in other words, the polynomial

$$P_n(h) = f(a) + f'(a)h + \frac{1}{2!}f''(a)h^2 + \dots + \frac{1}{n!}f^{(n)}(a)h^n \quad (3)$$

satisfies the relation (2) for $h \rightarrow 0$ and thus solves our problem. Assuming that

$$f(a+h) - P_n(h) = \varphi(h),$$

we must therefore show that

$$\frac{\varphi(h)}{h^n} \rightarrow 0 \quad (h \rightarrow 0).$$

But a simple calculation gives us *)

$$\varphi(h) = f(a+h) - f(a) - hf'(a) - \frac{h^2}{2!}f''(a) - \dots - \frac{h^n}{n!}f^{(n)}(a),$$

$$\varphi'(h) = f'(a+h) - f'(a) - hf''(a) - \dots - \frac{h^{n-1}}{(n-1)!}f^{(n)}(a),$$

$$\varphi''(h) = f''(a+h) - f''(a) - hf'''(a) - \dots - \frac{h^{n-2}}{(n-2)!}f^{(n)}(a),$$

.....

$$\varphi^{(n-2)}(h) = f^{(n-2)}(a+h) - f^{(n-2)}(a) - hf^{(n-1)}(a) - \frac{h^2}{2!}f^{(n)}(a),$$

$$\varphi^{(n-1)}(h) = f^{(n-1)}(a+h) - f^{(n-1)}(a) - hf^{(n)}(a),$$

hence

$$\varphi(0) = \varphi'(0) = \varphi''(0) = \dots = \varphi^{(n-2)}(0) = 0.$$

*) It evidently follows from our assumption on existence of $f^{(n)}(a)$ that $f^{(n-1)}(a+h)$ exists for a sufficiently small $|h|$ and therefore also $f^{(n-2)}(a+h), \dots, f'(a+h)$.

On the other hand, the function h^n vanishes together with its derivatives up to the order $n - 2$ inclusively even for $h \rightarrow 0$; the derivative of order $n - 1$ of this function is equal to $n! h$. Hence the application of l'Hospital's rule gives

$$\lim_{h \rightarrow 0} \frac{\varphi(h)}{h^n} = \lim_{h \rightarrow 0} \frac{\varphi^{(n-1)}(h)}{n! h}, \quad (4)$$

provided the limit on the right-hand side of this equation exists. But

$$\frac{\varphi^{(n-1)}(h)}{n! h} = \frac{1}{n!} \left\{ \frac{f^{(n-1)}(a+h) - f^{(n-1)}(a)}{h} - f^{(n)}(a) \right\}, \quad (5)$$

and since by definition

$$f^{(n)}(a) = \lim_{h \rightarrow 0} \frac{f^{(n-1)}(a+h) - f^{(n-1)}(a)}{h},$$

therefore the right-hand side, and hence also the left-hand side of the equation (5), tends to zero for $h \rightarrow 0$; it therefore follows from (4) that

$$\frac{\varphi(h)}{h^n} \rightarrow 0 \quad (h \rightarrow 0),$$

which proves our statement.

The formula (T) which we have thus established by assuming only the existence of $f^{(n)}(a)$ is usually known as *Taylor's formula*. This is one of the most important formulae of mathematical analysis and has a great number of theoretical and practical applications. It is sometimes convenient to write it down with the help of other symbols: Let us agree to write x in place of $a + h$; in this case $h = x - a$, and formula (T) becomes

$$\begin{aligned} f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots \\ \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + o[(x-a)^n]. \end{aligned}$$

In a particular case when $a = 0$, we obtain the so-called *Maclaurin-formula*

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + o(x^n),$$

which approximately represents the function $f(x)$ in the form of a polynomial in powers of x for small absolute values of x .

We have thus found that Taylor's polynomial (3) solves our problem of the approximate expression of $f(a+h)$ in the form of a polynomial of the n th degree, *i.e.* it satisfies the relation (2). We will now show that this solution of the problem is unique, *i.e.* no other polynomial $Q_n(h)$ of a degree not higher than n exists, for which we can also have

$$f(a+h) - Q_n(h) = o(h^n) \quad (6)$$

for $h \rightarrow 0$. In fact, if such a polynomial $Q_n(h)$ exists, then it would follow from (2) and (6) that

$$P_n(h) - Q_n(h) = o(h^n)$$

for $h \rightarrow 0$; but $P_n(h) - Q_n(h) = \beta_0 + \beta_1 h + \dots + \beta_n h^n$ is a polynomial of degree higher than n ; let β_k be the first of the numbers $\beta_0, \beta_1, \dots, \beta_n$ other than zero; then we have:

$$P_n(h) - Q_n(h) = \beta_k h^k + \beta_{k+1} h^{k+1} + \dots + \beta_n h^n = o(h^n) = o(h^k),$$

since $k \leq n$; but this means that for $h \rightarrow 0$

$$\frac{\beta_k h^k + \beta_{k+1} h^{k+1} + \dots + \beta_n h^n}{h^k} = \beta_k + \beta_{k+1} h + \dots + \beta_n h^{n-k} \rightarrow 0$$

for $h \rightarrow 0$, which is impossible for the limit of the left-hand side is evidently equal to $\beta_k \neq 0$ for $h \rightarrow 0$. This proves uniqueness of the solution of our problem.

§ 39. The last term in Taylor's formula

Taylor's formula gives us an expression $o(h^n)$ for the difference between the function $f(a+h)$ and the polynomial $P_n(h)$ (*i.e.* for the error in the approximate expression of $f(a+h)$ in terms of a polynomial); we know that this describes the *character of change* of the difference $f(a+h) - P_n(h)$ for $h \rightarrow 0$, but it tells us nothing about the *value* of this difference, *i.e.* for example, how small this difference is for a given value of h . It is obvious that in definite calculations when we replace the expression $f(a+h)$ by a polynomial $P_n(h)$, we need to know the magnitude of error caused by this substitution for the given values of a and h with which we are, in fact, dealing. We must therefore try to find a method for the assessment of this error and not be satisfied by the mere indication of the character of this change given by Taylor's formula. In other words: Taylor's formula gives us only the characteristics of the limiting behaviour of error in question but we want to know how to assess this error for the given definite values of a and h .

With this view let us write formula (T) in the form

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \dots \\ \dots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a) + R_n(h), \quad (1)$$

where we have assumed that

$$R_n(h) = \frac{h^n}{n!}f^{(n)}(a) + o(h^n).$$

The term $R_n(h)$ is called the *last term* in Taylor's formula.

Let a denote an arbitrary (not necessarily integral) positive number. For the sake of brevity we denote $a+h$ by b and consider the function

$$\varphi(x) = f(x) + (b-x)f'(x) + \frac{(b-x)^2}{2!}f''(x) + \dots \\ \dots + \frac{(b-x)^{n-1}}{(n-1)!}f^{(n-1)}(x) + \frac{R_n(h)}{(b-a)^q}(b-x)^q. \quad (2)$$

We have so far assumed the existence of the function $f^{(n)}(x)$ only for $x=a$; we shall now have to strengthen somewhat this assumption and assume that $f^{(n)}(x)$ exists for all interior points of the interval (a, b) . This evidently means that in these circumstances the function $\varphi(x)$ will be differentiable at every interior point. On differentiating we find that ($a < x < b$):

$$\begin{aligned} \varphi'(x) &= f'(x) + (b-x)f''(x) - f'(x) + \\ &+ \frac{(b-x)^2}{2!}f'''(x) - (b-x)f''(x) + \dots + \frac{(b-x)^{n-1}}{(n-1)!}f^{(n)}(x) - \\ &- \frac{(b-x)^{n-2}}{(n-2)!}f^{(n-1)}(x) + \frac{R_n(h)}{(b-a)^q}q(b-x)^{q-1} = \\ &= \frac{(b-x)^{n-1}}{(n-1)!}f^{(n)}(x) - \frac{R_n(h)}{(b-a)^q}q(b-x)^{q-1}. \end{aligned}$$

Further, we evidently have $\varphi(b) = f(b)$, and it can readily be seen from (1) that $\varphi(a) = f(a+h) = f(b)$ also. We can therefore apply Rolle's theorem to the function $\varphi(x)$ [in the interval (a, b)],

$\varphi'(c)=0$ at a point c situated between a and $b = a + h$. We can evidently assume that $c = a + \theta h$, where $0 < \theta < 1$; then we have

$$b - c = a + h - c = (1 - \theta) h,$$

and we find that

$$\begin{aligned}\varphi'(c) &= \frac{(b - c)^{n-1}}{(n - 1)!} f^{(n)}(c) - \frac{R_n(h)}{(b - a)^q} q(b - c)^{q-1} \\ &= \frac{(1 - \theta)^{n-1}}{(n - 1)!} h^{n-1} f^{(n)}(a + \theta h) - q R_n(h) \frac{(1 - \theta)^{q-1}}{h} = 0,\end{aligned}$$

hence

$$R_n(h) = \frac{h^n (1 - \theta)^{n-q}}{q(n - 1)!} f^{(n)}(a + \theta h).$$

This expression for the last term in Taylor's formula is very versatile owing to the presence of the parameter q which we can be given any arbitrary positive value. Evidently the problem as to which of these values gives $R_n(h)$ the most convenient form depends on the form of the function $f(x)$. However, in majority of cases it is most convenient to assume that $q = n$ so that $R_n(h)$ becomes

$$R_n(h) = \frac{h^n}{n!} f^{(n)}(a + \theta h). \quad (3)$$

Formula (1) then becomes

$$\begin{aligned}f(a + h) &= f(a) + h f'(a) + \frac{h^2}{2!} f''(a) + \dots \\ &\dots + \frac{h^{n-1}}{(n - 1)!} f^{(n-1)}(a) + \frac{h^n}{n!} f^{(n)}(a + \theta h) \quad (4);\end{aligned}$$

and represents a typical mean value theorem; for $n = 1$ it becomes Lagrange's theorem

$$f(a + h) = f(a) + h f'(a + \theta h),$$

and it does, in fact, represent a generalisation of that theorem. The form (3) for the last term in Taylor's formula was also introduced by Lagrange and is usually known by his name.

Among other forms of the last term in use let us note one more form which can be obtained from the general formula when $q = 1$:

$$R_n(h) = \frac{h^n (1 - \theta)^{n-1}}{(n - 1)!} f^{(n)}(a + \theta h)$$

(this is known as *Cauchy's form*).

Having obtained one or other expression for the last term in Taylor's formula we are now able to assess precisely the degree of accuracy given by this formula. In order to illustrate how this is done we shall now apply Taylor's formula to some simple elementary functions.

Example 1. $f(x) = e^x$, $a = 0$; it is convenient to write x instead of h ; formula (4) gives us $f^{(k)}(0) = 1$, $k = 1, 2, \dots$, since $f^{(k)}(x) = e^x$,

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^{n-1}}{(n-1)!} + \frac{x^n}{n!} e^{\theta x} \quad (0 < \theta < 1).$$

When $0 \leq x \leq 1$, say, then the last term of this formula does not exceed

$$\frac{x^n e}{n!} \leq \frac{e}{n!}$$

and, as n increases, it decreases rapidly even for small values of x ; in particular, when $x = 1$, we obtain the formula

$$e = 1 + 1 + \frac{1}{2!} + \dots + \frac{1}{(n-1)!} + \frac{1}{n!} e^{\theta}$$

which enables us to evaluate approximately the value of the number e with a high degree of accuracy, since the last term does not exceed $e/n!$ and, as we have said above, it decreases rapidly as n increases.

Example 2. $f(x) = \sin x$, $a = 0$; it can be readily seen that the numbers $f(0), f'(0), f''(0), \dots$ form a periodical sequence $0, 1, 0, -1, 0, 1, 0, -1, \dots$. Hence, formula (4) gives for $a = 0$, $h = x$ and odd $n = 2k + 1$,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!} + (-1)^k \frac{x^{2k+1}}{(2k+1)!} \cos \theta x,$$

since $f^{(2k+1)}(x) = (-1)^k \cos x$. A similar calculation for $f(x) = \cos x$ gives:

$$\begin{aligned} \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots + (-1)^{k-1} \frac{x^{2k-2}}{(2k-2)!} + \\ &\quad + (-1)^k \frac{x^{2k}}{(2k)!} \cos \theta x. \end{aligned}$$

In both expansions $|\cos \theta x| \leq 1$ and the last term whose absolute value does not exceed $|x|^{2k+1}/(2k+1)!$ or $|x|^{2k}/(2k)!$ decreases very rapidly as k increases, particularly for small values of $|x|$.

Example 3. $f(x) = \ln x$, $a = 1$; in this case

$f(x) = x^{-1}$, $f''(x) = -x^{-2}$, ..., $f^{(n)}(x) = (-1)^{n-1} (n-1)! x^{-n}$,
and therefore

$f(1) = 0$, $f'(1) = 1$, $f''(1) = -1$, ..., $f^{(n)}(1) = (-1)^{n-1} (n-1)!$,

so that formula (4) gives for $a = 1$ and $h = x$:

$$\begin{aligned} \ln(1+x) = & x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{n-2} \frac{x^{n-1}}{n-1} + \\ & + (-1)^{n-1} \frac{x^n}{n} (1+\theta x)^{-n}. \end{aligned}$$

When $0 < x \leq 1$, the absolute value of the last term is smaller than $1/n$ and therefore tends to zero when n increases, although not as rapidly as in the previous examples. But when $-1 < x < 0$, the factor $(1+\theta x)^{-n}$ is infinitely large and we cannot judge the order of its growth, for θ is unknown. In this case Cauchy's form for the last term is more convenient

$$R_n(x) = \frac{x^n (1-\theta)^{n-1}}{(n-1)!} f^{(n)}(1+\theta x) = (-1)^{n-1} \frac{x^n}{1+\theta x} \left(\frac{1-\theta}{1+\theta x} \right)^{n-1};$$

when $-1 < x < 0$, we have $0 < (1-\theta)/(1+\theta x) < 1$; the absolute value of the last term is smaller than $|x|^n/(1-|x|)$ and tends to zero for $n \rightarrow \infty$.

In accordance with the problem stated at the beginning of this paragraph we were trying to assess the quantity $R_n(h)$ for the given values of a , h and n , i.e. we were trying to assess the magnitude of the error caused by replacing $f(a+h)$ by a polynomial of the n th degree. In practice, we often have to solve problems which are in a sense, opposed to the above problems; thus the permissible limit of error Δ is often given at the start; the problem usually involves two questions: either — within what limits of variation of h can we guarantee, for a given n , that $|R_n(h)|$ will not exceed Δ or, conversely, how large should the number n be taken for the given limits of variation of h so as to achieve this aim. We will show below how such problems can be solved on the basis of the most widely used form (3) for the last term.

Let us assume that we are interested in the values of the function $f(x)$ in interval $(a-l \leq x \leq a+l)$. If we denote by

$M^{(n)}$ the maximum value of the function $|f^{(n)}(x)|$ in that interval, then as a result of (3) we have $|R_n(h)| \leq M^{(n)} |h|^n / n!$ for $|h| \leq 1$; hence in order to guarantee the required assessment $|R_n(h)| < \Delta$ it is sufficient that :

$$\frac{M^{(n)} |h|^n}{n!} < \Delta,$$

or

$$|h| < \left(\frac{\Delta n!}{M^{(n)}} \right)^{\frac{1}{n}}. \quad (*)$$

Hence if $|h|$ is smaller than the smallest of the numbers l and $(\Delta n! / M^{(n)})^{1/n}$, then the inequality $|R_n(h)| < \Delta$ can be guaranteed. If the limit of the error Δ and the number l which determines the amplitude of variation are given for the given calculation, as it frequently happens, then we must take n sufficiently large so that

$$\left\{ \frac{\Delta n!}{M^{(n)}} \right\}^{\frac{1}{n}} \geq l.$$

In that case, the inequality (*) will be satisfied for $|h| < l$ and hence also the required inequality $|R_n(h)| < \Delta$.

If, for example, $f(x) = \sin x$ or $f(x) = \cos x$, cf. the above example 2, then it is very easy and natural to assume in calculations that $a = 0$, $l = \pi/4$, for by knowing the values of $\sin x$ and $\cos x$ in the interval $(0, \pi/4)$ we can find without further calculations the values of these functions for every x . Since $|f^{(2k+1)}(x\theta)| = |\cos \theta x| \leq 1$ for $f(x) = \sin x$, therefore we can assume that $M^{(2k+1)} = 1$. Let the required degree of accuracy be $\Delta = 0.0001$. We should then have :

$$0.0001 \cdot (2k+1)! \geq \left(\frac{\pi}{4} \right)^{2k+1},$$

which, as can readily be calculated, happens when $k \geq 3$. Hence the approximate formula

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

gives the value of the function $\sin x$ in the interval $|x| \leq \pi/4$ with an error not exceeding 0.0001. The calculation is similar for $f(x) = \cos x$.

CHAPTER X

APPLICATION OF DIFFERENTIAL CALCULUS TO ANALYSIS OF FUNCTIONS

§ 40. Increasing and decreasing functions

The true meaning of a derivative which leads us to its general definition implies that the absolute value $|y'| = |f'(x)|$ of the derivative determines the rate of change of the function $y = f(x)$ in relation to the independent variable x ; hence by knowing the derivative of the given function we can, in the majority of cases, directly find the rate of change of the function in a given interval. In order to appreciate significance of this information let us consider the following example. Both functions $y = x^2$ and $z = \ln x$ increase together with x , for $x > 0$. To find the rate of this increase consider their derivatives

$$y' = 2x, \quad z' = \frac{1}{x};$$

we can see that as x increases, y' increases continuously while z' decreases continuously; this means that as x increases, the function $y = x^2$ increases at an increasing rate while the rate of increase of the function $z = \ln x$ decreases; thus although both functions increase as x increases, their respective rates of increase show the above differences; we are able to detect these differences by simply looking at their derivatives y' and z' . This difference in their behaviour can also be readily detected by looking at the graphs of these functions (Fig. 20), but it is one of the advantages of the derivative that it does not necessitate construction of a graph for the given function in order to find its rate of change.

On the other hand we have already seen that the *sign* of the derivative determines the *direction of the change* of the function: a positive derivative implies increase and a negative derivative decrease of

a function (both are related to the increase of the independent variable). We must now state this problem more precisely.

Let us agree to say that the function $y = f(x)$ defined in the interval (a, b) is non-decreasing in that interval if we always have $f(x_2) \geq f(x_1)$ for $a \leq x_1 < x_2 \leq b$ (i.e. if in that interval y cannot decrease as x increases); if, however, the exact inequality $f(x_2) > f(x_1)$ holds for $a \leq x_1 < x_2 \leq b$, then we shall call the function $y = f(x)$ an increasing function in the interval (a, b) . Similarly, if the signs of the inequalities for $f(x_1)$ and $f(x_2)$ are interchanged, they are defined as *non-increasing* and *decreasing* functions, respectively, in that interval. It is obvious that every increasing function is also non-decreasing, but the converse is not true; similarly every decreasing function is also non-increasing, but the converse is not true.

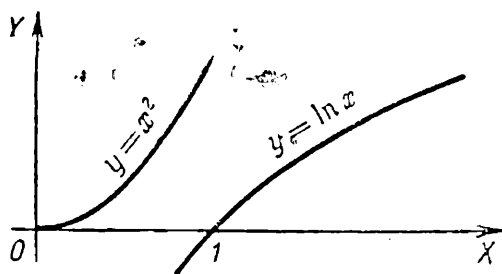


Fig. 20

The relationship between the sign of the derivative and the direction of change of the function is expressed by the following propositions.

Theorem 1. *In order that the function $f(x)$, differentiable at every point in the interval (a, b) , should be non-decreasing in that interval, it is necessary and sufficient that*

$$f'(x) \geq 0 \quad (a \leq x \leq b).$$

Proof. 1) If the function $f(x)$ is non-decreasing in (a, b) for $a \leq x < x + h \leq b$, then

$$f(x + h) - f(x) \geq 0,$$

and therefore also

$$\frac{f(x + h) - f(x)}{h} \geq 0;$$

it follows from corollary 2, theorem 2 § 10 that we also have

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h} \geq 0.$$

2) If $f'(x) \geq 0$ ($a < x \leq b$), then it follows from the theorem on finite increments that for $a \leq x_1 < x_2 \leq b$

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) \geq 0$$

(here c is a point between x_1 and x_2 , and hence also between a and b). But this means that the function $f(x)$ is non-decreasing in the interval (a, b) . Theorem 1 is thus proved.

Theorem 1 evidently remains valid when the word “non-decreasing” is replaced by “non-increasing” and instead of writing $f'(x) \geq 0$ we write $f'(x) \leq 0$. To prove this, it is sufficient to apply the statement of theorem 1 to the function $-f(x)$.

Theorem 2. *If $f'(x) > 0$ ($a \leq x \leq b$), then the function $f(x)$ increases in the interval (a, b) .*

Proof. The theorem on finite increments gives us for $a \leq x_1 < x_2 \leq b$

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) > 0,$$

since $x_1 < c < x_2$ and therefore also $a < c < b$, hence $f'(c) > 0$.

The property $f'(x) > 0$ ($a \leq x \leq b$) is therefore *sufficient* in order that the function $f(x)$ should increase in the interval (a, b) , but it is not the *necessary* condition; the converse of theorem 2 is incorrect; $f(x_2) > f(x_1)$ ($a \leq x_1 < x_2 \leq b$) implies (as a result of theorem 1) only that $f'(x) \geq 0$ ($a \leq x \leq b$) but not that $f'(x) > 0$ ($a \leq x \leq b$); this is illustrated by the function $f(x) = x^3$, which increases on the whole number line ($-\infty < x < +\infty$) whereas $f'(x) = 3x^2 = 0$ for $x = 0$; Fig. 21 clearly illustrates this phenomenon; the curve $y = x^3$ increases continuously from left to right and at the same time has a horizontal tangent at $x = 0$.

It is self-evident that we always have $f'(x) < 0$ for $a \leq x \leq b$ and the function $f(x)$ decreases in the interval (a, b) ; the converse statement is, in this case, not true.

Example 1. The function $y = x^3 - 6x^2 + 9x + 2$ has the derivative

$$y' = 3x^2 - 12x + 9 = 3(x - 1)(x - 3);$$

the brackets $(x - 1)$ and $(x - 3)$ have opposite signs for $1 < x < 3$ and have the same sign for $x < 1$ and $x > 3$; therefore

$$y' > 0 \quad (x < 1 \text{ or } x > 3) \text{ and } y' < 0 \quad (1 < x < 3);$$

the function y increases for $x < 1$ and $x > 3$ and decreases for $1 < x < 3$. A simple calculation gives us:

$$y = 6 \quad (x = 1), \quad y = 2 \quad (x = 3),$$

and, on the other hand, it is evident that

$$y \rightarrow -\infty \quad (x \rightarrow -\infty), \quad y \rightarrow +\infty \quad (x \rightarrow +\infty),$$

therefore the sample graph of the function y can be fairly drawn on the basis of this short analysis (Fig. 22).

Example 2. The function $y = e^x - x - 1$ has the derivative

$$y' = e^x - 1,$$

so that $y' > 0$ for $x > 0$ and $y' < 0$ for $x < 0$. The function y increases for $x > 0$ and decreases for $x < 0$; it is equal to zero for $x = 0$ and therefore it must be positive for all other values of x ; this proves the important inequality

$$e^x \geq 1 + x,$$

which holds for every real x , but the sign of equality holds only for $x = 0$.

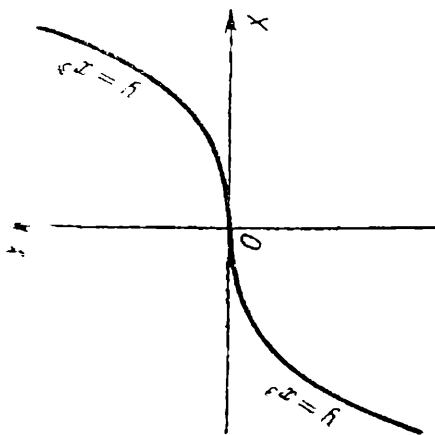


Fig. 21.

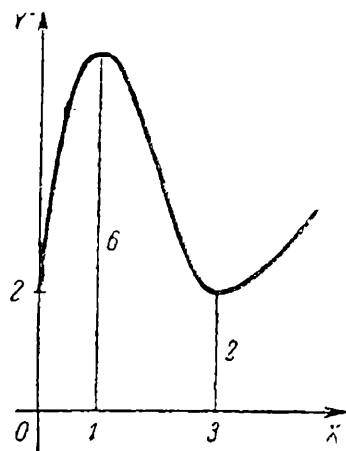


Fig. 22.

For other useful exercises cf. problem book by B.P. Demidovich, § II, Nos. 320-325, 332, 339.

§ 41. Extrema

Let $y = f(x)$ be a function differentiable at every point in the interval (a, b) . We know that this function is continuous in that interval and (theorem 2 § 23) attains its maximum and minimum values in that interval. In view of practical applications it would be interesting to know, for what values of the independent variable the function takes its maximum (or minimum) values. Thus $y = f(x)$

can measure efficiency of a plant which depends on the choice of x , a choice which can be arbitrary within the limits (a, b) . In such cases we would evidently choose an x in that interval such that y assumes its maximum values (and, of course, we shall also be interested in this maximum value). We shall now learn to appreciate significance of methods of differential calculus in such cases.

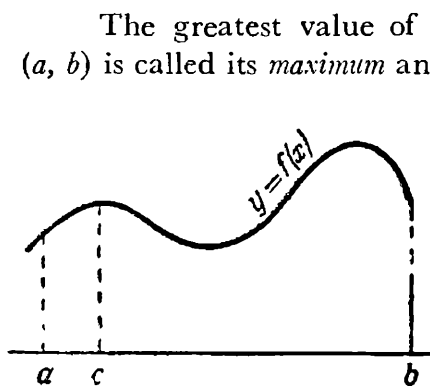


Fig. 23.

The greatest value of the function $y = f(x)$ in the interval (a, b) is called its *maximum* and the lowest value *minimum*; when we want to say “maximum or minimum”, we say more briefly “extrema” or “extreme values” (*i.e.* border values). If we speak of extrema of the function *in the whole interval* (a, b) , we speak of the “absolute” extremes (maximum or minimum) of the function. The concept of *local* extreme (as applied to a given spot) implies as follows: the function $f(x)$ has a *local maximum* at the point c ($a < c < b$) if its value at the point c is greatest in comparison to its values at all points sufficiently close to c , *i.e.* if a number δ exists such that

$$f(c + h) \leq f(c)$$

for all h , for which $|h| \leq \delta$. The *local minimum* is defined similarly. As a result of the condition $a < c < b$ the local extremum is only defined for *interior* points in the interval (a, b) .

Fig. 23 shows the difference between the absolute and local extrema: the function $y = f(x)$ represented in this graph has its local maximum at the point c which, however, is not its absolute maximum, for at a certain distance from the point c the function assumes a greater value than $f(c)$.

It is obvious that if the function assumes its greatest value at an interior point c in the interval (a, b) , then the absolute maximum will also be its local maximum.

Hence in order to find the absolute maximum (or minimum) of the given function in the given interval we must proceed as follows:

1) find the local maximum (or minimum) of the function in the given interval;

2) add to the values of the function at points of its local maximum (or minimum) its values at the ends of the given interval and choose from all these values the greatest (smallest) value *).

The second problem presents no difficulties if, as is usually the case in practical problems, the function has only a finite (and usually small) number of extrema. Thus the difficulty is concentrated in the first problem which we can now solve with the help of methods of differential calculus (provided the function is differentiable).

Let the function $y = f(x)$ have a local maximum at an interior point $x = c$ in the interval (a, b) and let it be differentiable at that point. On the basis of the lemma in § 36 we can therefore maintain that $f'(c) = 0$; we evidently also arrive at the same conclusion when $f(x)$ has a local minimum at the point c . Thus if the function $f(x)$ is differentiable at every point in the interval (a, b) , then all its local extrema, if they exist, will be found among the roots of the equation

$$f'(x) = 0. \quad (1)$$

We must therefore begin the solution of our problem by finding all roots of this equation between a and b . The roots of the equation (1) are usually called *stationary points* of the function $f(x)$; this term is quite intelligible: the rate of change of the function $f(x)$ at such points is equal to zero; as x passes through such a point, $f(x)$ changes very slowly and its value is particularly stable.

We must therefore find all stationary points of the function $f(x)$ in the interval (a, b) and all the required local extrema will then be found among those points. Let α be one such stationary point; we must find out whether it gives a local extremum and, if so, the type of this extremum, *i.e.* a maximum or a minimum. Let us now assume that the function $f(x)$ has a derivative of the first order and several more derivatives of higher orders at the point α ; let us assume that in general

$$f'(\alpha) = f''(\alpha) = \dots = f^{(n-1)}(\alpha) = 0,$$

but $f^{(n)}(\alpha) \neq 0$ (if $f''(\alpha) \neq 0$, then $n = 2$). In this case we evidently obtain from Taylor's formula (T) § 38:

$$f(\alpha + h) - f(\alpha) = h^n \frac{f^{(n)}(\alpha)}{n!} + o(h^n),$$

*) If the function has no derivatives at certain points in given interval, then all such points must be added to the ends of the interval. The reader should consider the example $y = |x|$ ($-1 \leq x \leq 1$).

and in order to solve our problem we must analyse the sign of the difference $f(\alpha + h) - f(\alpha)$ for sufficiently small values of $|h|$. Since the second term on the right-hand side of the last equation is infinitely small as compared to the first term for $h \rightarrow 0$, the sign of the whole right-hand (and also left-hand) side of this equation will, for sufficiently small values of $|h|$ coincide with the sign of its first term which we must therefore analyse.

If n is an even number, then $h^n > 0$ and the sign of the expressions $h^n f^{(n)}(\alpha) / n!$ coincides with the sign of $f^{(n)}(\alpha)$ (this also means that it is independent of h); if $f^{(n)}(\alpha) > 0$, then we have for a sufficiently small $|h|$;

$$f(\alpha + h) - f(\alpha) > 0,$$

i.e. the function $f(x)$ has its *minimum* at the point α ; on the other hand if $f^{(n)}(\alpha) < 0$, then for all sufficiently small values of $|h|$ we have :

$$f(\alpha + h) - f(\alpha) < 0,$$

i.e. the function $f(x)$ has its *maximum* at the point α .

If n is odd, then h^n , and therefore also the expression $h^n f^{(n)}(\alpha) / n!$ changes its sign when the sign of h changes and therefore, provided $|h|$ is sufficiently small, the difference $f(\alpha + h) - f(\alpha)$ will have one sign for positive h and the opposite sign for negative h ; this evidently means that the function $f(x)$ can neither have a maximum nor a minimum at the point α (an example of this kind is given by $f(x) = x^3$ for $x = 0$: $f'(0) = f'''(0) = 0, f''(0) \neq 0$ (cf. Fig. 21 § 40, where the point $x = 0$ gives a typical example of a stationary point without a local extremum).

We thus obtain (on the assumption that the function $f(x)$ can be differentiated a sufficient number of times) a fully defined method for the analysis of the character of every stationary point α : Let among a sequence of derivatives $f'(\alpha), f''(\alpha), \dots$ the derivative $f^{(n)}(\alpha)$ be the first which is not equal to zero; then 1) if n is odd, function $f(x)$ has neither a maximum nor a minimum at the point α ; 2) if n is even, a local extremum exists at the point α which will be the local minimum for $f^{(n)}(\alpha) > 0$ and the local maximum for $f^{(n)}(\alpha) < 0$.

In particular, we have a local minimum for $f'(\alpha) = 0, f''(\alpha) < 0$ and a local maximum for $f'(\alpha) = 0, f''(\alpha) > 0$ we must analyse the derivatives of higher orders for $f'(\alpha) = f''(\alpha) = 0$.

The above method for the analysis of stationary points may be inapplicable only when the function $f(x)$ has no derivatives of sufficiently high orders at the point α or when its derivatives of all orders are equal to zero. It is interesting to note that the latter case may, in fact, take place (and we do, of course, exclude the trivial case when $f(x)$ is simply a constant in a certain neighbourhood of α). The function

$$y = f(x) = \begin{cases} e^{-1/x^2} & (x \neq 0), \\ 0 & (x = 0) \end{cases}$$

has a stationary point for $x = 0$, where

$$y' = y'' = \dots = y^{(n)} = \dots = 0,$$

whereas the behaviour of this function is very simple in the neighbourhood of the point $x = 0$ (cf. Fig. 24) and it differs very little from the behaviour of functions like $y = x^2$ or $y = x^4$; a difference in the behaviour of these functions can be detected only at a certain distance from $x = 0$.

The established method of analysis of stationary points is significant from a theoretical point of view which for its finality is, in practice, often replaced by simpler methods which are also more convenient insofar as they do not necessitate existence of derivatives of higher orders. If α is the stationary point of the function $f(x)$, i.e. if $f'(\alpha) = 0$, then in order to determine the character of this point it is sufficient in many cases to determine only the sign of the derivative $f'(x)$ in the immediate neighbourhood of the point α ; thus we always have $f'(x) < 0$ for $x < \alpha$ (provided $|x - \alpha|$ is sufficiently small) and we have $f'(x) > 0$ for $x > \alpha$ (and $|x - \alpha|$ sufficiently small); hence the function $f(x)$ decreases on the left of α and increases on the right of α : therefore at the point α it has its local minimum;

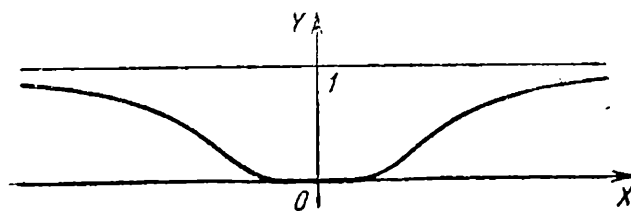


Fig. 24.

when the signs of $f'(x)$ are reversed, we obtain the local maximum; when $f'(x)$ maintains the same sign for all values of x sufficiently

close to α , then, as x passes through α , the function $f(x)$ either continues to increase or decrease but in neither case it can have a local extremum at the point α .

In spite of the simplicity of this method it must not be overrated. The determination of the sign of $f'(x)$ for *all* values of x sufficiently close to α is in many cases not easier but more difficult than the evaluation of several derivatives at the point α . Moreover, this method can only give results if, for example, $f'(x)$ has on the right of α one and the same sign for all values of x sufficiently close to α ; this may, of course, not be so; it may happen, for example, that as $x \rightarrow \alpha + 0$, the derivative $f'(x)$ changes its sign an infinite number of times; if this is the case, the method described above is inapplicable in principle.

Example 1. Find the absolute maximum and the absolute minimum of the function

$$f(x) = x^3 - 6x^2 + 9x + 2$$

in the interval $(0, 4)$. While investigating this function in § 40 we have found that it has two stationary points: $x = 1$ and $x = 3$; the first of these gives the local maximum and the second the local minimum. On adding to them the ends of the given line, we find that the following points can only be absolute extrema for $f(x)$: 0, 1, 3 and 4. We have

$$f(0) = 2, \quad f(1) = 6, \quad f(3) = 2, \quad f(4) = 6;$$

hence the function $f(x)$ has two absolute maxima (for $x = 1$ and $x = 4$) and two absolute minima (for $x = 0$ and $x = 3$) in the interval $(0, 4)$.

Example 2. Find all local extrema of the function

$$f(x) = \sinh x - x = \frac{e^x - e^{-x}}{2} - x.$$

$$f'(x) = \cosh x - 1 = \frac{e^x + e^{-x}}{2} - 1,$$

and we can immediately see that $x = 0$ is a stationary point. [$f'(0) = 0$]; further

$$\begin{aligned} f''(x) &= \sinh x, & f''(0) &= 0, \\ f'''(x) &= \cosh x, & f'''(0) &= 1; \end{aligned}$$

hence the first non-vanishing derivative is of an odd (third) order and therefore the function $f(x)$ has no local extremum at the stationary point $x = 0$. It remains to be shown that no other stationary points exist. We can directly see from the expression obtained for $f''(x)$ that

$$f''(x) \begin{cases} < 0 & (x < 0), \\ > 0 & (x > 0), \end{cases}$$

and therefore $f'(x)$ decreases for $x < 0$ and increases for $x > 0$; since $f'(x) = 0$, therefore $f'(x) > 0$ for all $x \neq 0$, i.e. apart from the point $x = 0$ the function $f(x)$ has no other stationary points (the graph of this function is similar to the graph of the function $y = x^3$, cf. Fig. 21 § 40). Hence the function $f(x)$ has no local extrema.

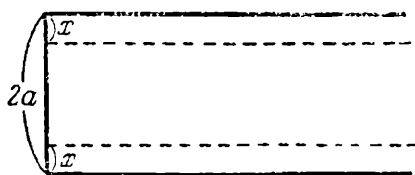


Fig. 25.

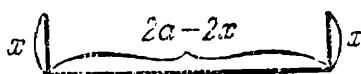


Fig. 26.

Example 3. From a rectangular section of a tin sheet $2a$ wide, strips of width x are bent on each side (Fig. 25) so as to make (an open from the top) trough whose cross-section is shown in Fig. 26. How wide should be the bent sides x so that the trough obtained should have maximum capacity?

It is evident that the length of the trough is not involved in the solution of this problem; the capacity of the trough is proportional to its cross-sectional area which is equal to $2x(a - x)$. We must therefore find the absolute maximum of the function

$$f(x) = 2ax - 2x^2$$

in the interval $(0, a)$. We have:

$$f'(x) = 2a - 4x,$$

and the only stationary point is $x = a/2$; since $f''(x) = -4 < 0$ (for every x and, in particular, for $x = a/2$), $f(x)$ has a local maximum for $x = a/2$:

$$f\left(\frac{a}{2}\right) = \frac{1}{2} a^2.$$

This will also be the absolute maximum, since $f(0) = f(a) = 0$. Hence it is most convenient to bend the edges so that their width should comprise of one quarter of the width of the given strip.

Further exercises can be found in any problem book on mathematical analysis. We can recommend exercises in B.P. Demidovich's Problem Book Section II, Nos. 436-444, 448, 452-466, 468-472, 539, 541, 542, 546, 552, 558.

CHAPTER XI

INVERSE OF DIFFERENTIATION

§ 42. Concept of primitives

If the law of motion of a body is given by an equation of the type

$$s = f(t),$$

where t is time and s is distance travelled by the body, then by differentiating the function $f(t)$ we can find the instantaneous velocity

$$v = f'(t)$$

of this motion at the given instant. However, the converse problem is more frequently met in mechanics: we are given velocity $v = v(t)$ at a moment of time t and we are required to find the law of motion of the body, *i.e.* the relationship between the distance travelled and time taken by the body. How can we solve this problem? We know that the instantaneous velocity $v = v(t)$ is a derivative of the function $s = f(t)$ which expresses the required law of motion of the body. Thus the derivative $f'(t) = v(t)$ of an unknown function $f(t)$ is given and we must find this function. This problem is evidently inverse of the fundamental problem of differential calculus: there we had to find the derivative of a given function whereas here we must find the initial function from its given derivative.

We are given, for example, that at the instant t the velocity of a moving body is equal to

$$v = at,$$

where a is a constant. How can we find the law of motion? To solve this problem we must find a function whose derivative coincides with at . We know one such function

$$\frac{at^2}{2}.$$

Can we therefore say that the required law of motion will be

$$s = \frac{at^2}{2}?$$

This would evidently be premature, for apart from the function $at^2/2$ other functions may exist whose derivatives are also equal to at ; and if this is so and we have no other information, we cannot say which of these functions will give the required law of motion. It can readily be seen that other functions of this type exist: the function

$$s = \frac{at^2}{2} + b, \quad (1)$$

where b is an arbitrary constant, gives a derivative at ; we thus obtain a whole family of functions each of which can express equally the required law of motion. At the same time we do not even know whether the family (1) contains all functions which have the given derivative at , for other functions may exist outside the family (1).

The above problem can readily be generalised. As we know, the derivative $f'(x)$ of the given function $f(x)$ always expresses the rate of change of this function (in relation to the independent variable x). In many problems we are required to find a function whose rate of change (in relation to x) is given for every x . From mathematical point of view this type of problem always involves finding the unknown function from its given derivative, i.e. it is inverse of the fundamental problem of differential calculus. Let us now state this problem in its most general form; to do so, we shall introduce the necessary terminology and examine possibilities of finding a solution.

Thus we are given a function $f(x)$ in a segment (or on the whole number line); it is necessary to find the whole set of such functions $F(x)$ for which we have at every point on the given segment:

$$F'(x) = f(x).$$

We shall call such a function a *primitive* of the function $f(x)$ so that the concept of a derivative and a primitive are reciprocal *).

We cannot evidently forecast whether the given function $f(x)$ has primitives and, if it has, then how many, and how they are interdependent. However, we can establish certain facts in this connection

*) A primitive is also known as an *indefinite integral* of the given function; we are, however, not going to use that term.

from elementary considerations: to begin with, if $F(x)$ is one of the primitives of the given function $f(x)$, then every function of the family

$$F(x) + C, \quad (2)$$

where C is an arbitrary constant, will evidently also be a primitive of the function $f(x)$. Let us now show that no primitive of the function $f(x)$ exists outside the family (2). In fact, let $\varphi(x)$ be an arbitrary primitive of the function $f(x)$; we can find the difference $\varphi(x) - F(x)$; since the derivative of this difference is evidently equal to zero for every x , it follows from the last theorem in § 36 that $\varphi(x) - F(x)$ is a constant which we can denote by a . Hence

$$\varphi(x) = F(x) + a,$$

i.e. every primitive $\varphi(x)$ of the function $f(x)$ belongs to the family (2).

We thus deduce the following important result:

Theorem. *If the function $f(x)$ has a primitive $F(x)$, then it has an infinite number of primitives, all of which belong to the family (2).*

The importance of this result is self-evident: it shows that in order to find *all* primitives of the function $f(x)$ it is sufficient to find *any one* of them; if we succeed in finding one such primitive, then every other primitive is obtained from it by adding a constant. Hence the problem we tried to solve is simplified: we must find out whether the function $f(x)$ has at least one primitive and if so, we must find this primitive.

Finding primitive of the given function is known as *integration* of this function. We can say that integration involves transition from the derivative of a function to the function itself. If we regard this transition as an operation, we can say that integration is inverse of differentiation: if the given function is first differentiated and then integrated, then, by choosing a suitable constant C , we obtain the initial function by means of formula (2).

Let us now recall that we have agreed to understand by the term *differentiation* the finding of both the derivatives of the given function and its differential. The inverse operation, *viz.* integration, can therefore involve finding the function from its derivative as well as from its differential. The differential $dF(x)$ of the required function is equal to $F'(x) dx$; therefore finding the derivative or the differential is one and the same thing.

We obtain a primitive as a result of integration. Hence every differentiable function $F(x)$ is primitive of its derivative $F'(x)$ or of its differential $dF(x) = F'(x) dx$.

Integration is denoted by the symbol \int . Finding a primitive (integration) is an operation inverse of finding the differential (differentiation) which we denote by the symbol d . Hence the symbols d and \int express two inverse operations. If we subject the given function $F(x)$ to the operation d and then perform the operation \int , then, by choosing a suitable constant term, we return to the initial function $F(x)$:

$$\int dF(x) = F(x),$$

or, since $dF(x) = F'(x) dx$,

$$\int F'(x) dx = F(x).$$

If $F'(x) = f(x)$, then

$$F(x) = \int f(x) dx;$$

hence it can be seen from this formula that the function $F(x)$ is primitive of the function $f(x)$. By the way it is accepted to understand by the expression

$$\int f(x) dx$$

not a particular primitive but the *whole family* of primitives of the function $f(x)$; if $F(x)$ is one such primitive, then it is written as

$$\int f(x) dx = F(x) + C, \quad (3)$$

where C denotes an arbitrary constant, *i.e.* the so-called constant of "integration". It is evident that in view of its definition the equation (3) is equivalent to the equation

$$F'(x) = f(x).$$

The function $f(x)$ on the left-hand side of the equation (3) is known as the "integrand" and the product $f(x) dx$ as the "integrand expression".

Example 1. Since $d(x^3) = 3x^2 dx$, so

$$\int 3x^2 dx = x^3 + C.$$

Example 2. Since $d \tan x = \frac{dx}{\cos^2 x}$, so

$$\int \frac{dx}{\cos^2 x} = \tan x + C,$$

and so on.

It can be readily seen from these examples that a formula of the derivative (or differential) of an arbitrary function also gives us an integration formula only by reading it, as it were, from right to left. Keeping this in mind while looking at the table of derivatives of simple functions given at the end of § 29 we can draw the following conclusions :

1. $\int 0 \cdot dx = C$ (the primitive of zero is equal to an arbitrary constant).

2. $\int 1 \cdot dx = x + C$, and in general

$$\int a \, dx = ax + C,$$

where a is a constant.

3. For every $\alpha \neq -1$ and $x > 0$

$$\int x^\alpha \, dx = \frac{x^{\alpha+1}}{\alpha+1} + C,$$

and at the same time (if $x > 0$)

$$\int x^{-1} \, dx = \int \frac{dx}{x} = \ln x + C.$$

The following remark must be added to this statement. Owing to the fact that the function $\ln(-x)$ has the derivate $1/x$ for $x < 0$, we have

$$\int \frac{dx}{x} = \ln(-x) + C$$

for $x < 0$; hence we have the general formula

$$\int \frac{dx}{x} = \ln |x| + C$$

for $x > 0$ and $x < 0$.

$$4. \quad \int e^x dx = e^x + C,$$

and for every positive $a \neq 1$

$$\int a^x dx = \frac{a^x}{\ln a} + C.$$

5. For the polynomial $P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_n$ we have

$$\int P(x) dx = \frac{a_0 x^{n+1}}{n+1} + \frac{a_1 x^n}{n} + \dots + a_n x + C,$$

so that the primitive of a polynomial is always a polynomial one degree higher than the given polynomial.

$$6. \quad \int \sin x dx = -\cos x + C,$$

$$\int \cos x dx = \sin x + C,$$

$$\int \frac{dx}{\cos^2 x} = \tan x + C,$$

$$\int \frac{dx}{\sin^2 x} = -\cot x + C,$$

$$7. \quad \int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C = -\arccos x + C,$$

$$\int \frac{dx}{1+x^2} = \arctan x + C = -\operatorname{arccot} x + C.$$

$$8. \quad \int \sinh x dx = \cosh x + C,$$

$$\int \cosh x dx = \sinh x + C.$$

All these formulae can be verified by the same method: it is sufficient to show that the derivative of the right-hand side is equal to the integrand on the left-hand side; this follows in all cases from the corresponding formulae in the table of derivatives at the end of § 29.

We have thus learnt to find primitives of a series of simple functions. However, our knowledge in this direction is still very limited: we have only learnt to integrate functions which happen to be on the right-hand sides of differentiation formulae collected in that table. But these functions do not even include all simple elementary functions; they do not include functions like $\ln x$, $\arctan x$ and many other functions; and we have so far not met a function whose derivative is equal to $\ln x$ or $\arctan x$; therefore we are not only unable to find the primitives

$$\int \ln x \, dx \quad \text{or} \quad \int \arctan x \, dx$$

but also do not know whether they exist.

Integration is much more complicated and difficult than differentiation. This is first of all due to its nature. Finding a derivative of a given function is facilitated by the fact that the definition of differentiation itself has a "constructive" character; a derivative is simply defined as

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h},$$

i.e. as a series of definite operations which must be performed over the given functions; for example, having been asked to find the derivative of the function $\sin x$ we know in all details how to proceed in order to obtain a result; while trying to find a primitive, we are confronted with a totally different problem; we have no constructive element and we are given no method which would tell us how to find the primitive, or even how to begin. For example, when we are asked to find

$$\int \ln x \, dx$$

and we do not find $\ln x$ among the right-hand sides of the formulae in our table, we are unable to tackle the given problem and find its solution.

This position is further complicated by the fact that we do not even have a sequence of laws for integration which, in the case of differentiation, enable us to use the laws of differentiation of several functions in order to find derivatives of their various combinations, *e.g.* their sums, products, composite functions, etc. There are very few such rules available in the theory of integration and their

application is rather restricted. Nevertheless, the significance of these general methods of integration is very great, for they do eventually enable us to integrate a fairly large number of most general functions. In the following paragraph we shall consider several simple methods. It must also be stressed that in contrast to the general methods of differentiation which are used almost mechanically, the application of general methods of integration requires great skill—in every case one must learn to select the most suitable method and use it in the most convenient form. This skill can only be acquired after long practice.

The reader will find a large number of examples in Section III of the Problem Book by B.P. Demidovich.

The science which deals with integration of functions and properties of primitives is known as *integral calculus* and, together with differential calculus, comprises one of the most important branches of mathematical analysis.

§ 43. Simple general methods of integration

If $y = u_1 \pm u_2 \pm \dots \pm u_n$ is an algebraic sum of n functions of x and if $\int u_k dx$ exists for all k ($1 \leq k \leq n$), then $\int y dx$ also exists and

$$\int y dx = \int u_1 dx \pm \int u_2 dx \pm \dots \pm \int u_n dx. \quad (1)$$

This rule is often formulated in short as follows: “primitive of an algebraic sum is equal to the algebraic sum of primitives”. To prove this it is sufficient to show that derivative of the right-hand side of the equation (1) exists and is equal to y ; but this follows from the law of differentiation of an algebraic sum (law 3° § 29).

2. If u is a function of x and a is a constant and if $\int u dx$ exists, then $\int au dx$ also exists and

$$\int au dx = a \int u dx. \quad (2)$$

In short: “a constant factor can be taken outside the sign of integration”. To prove this it is sufficient to differentiate the right-hand side and use the theorem which states that a constant factor can be taken outside the symbol of differentiation.

In the two cases considered so far we find that the corresponding laws of differentiation can be fully used in an inverse form. But

as we shall now see, there are only these two cases; all other laws are only partially convertible and result in integration laws which can sometimes be very useful but not applicable in all cases.

We must finally note that we have omitted constants of integration on the right-hand sides in the relation (1) and (2); it is not necessary, for in both cases the symbol of the primitive stands on the right-hand side (or several such symbols); according to our agreement this symbol embraces the whole family of primitives and thus contains, as it were, an invisible arbitrary constant. This note also applies to a series of subsequent formulae.

3. Integration by parts. Let us now see what we obtain as a result of inverting the formula

$$(uv)' = uv' + vu',$$

which expresses the law of differentiation of a product of two functions. Integrating the above equation we have :

$$\int (uv)' dx = uv = \int (uv' + vu') dx;$$

applying the rule (1) to the right-hand side we obtain :

$$uv = \int uv' dx + \int vu' dx.$$

This formula contains two primitives; therefore by using this formula we are unable to find both primitives and we can only express either of them in terms of the other, for example

$$\int uv' dx = uv - \int vu' dx \quad (3)$$

What does this formula give us? If the functions u and v are known, we can express the primitive $\int uv' dx$ in terms of the known function uv and another primitive $\int vu' dx$, which has the same structure as the initial primitive. However, in relation to the form of the functions u and v one of these primitives may appear simpler than the other; if, for example, the second primitive (on the right-hand side) is simpler than the first, then formula (3) is undoubtedly useful, for we can replace the given primitive by another primitive of simpler form. It may sometimes happen that the primitive on the right-hand side belongs to the elementary table in § 29 or may be known on the basis of some preliminary considerations; in such cases the formula enables us to obtain the primitive $\int uv' dx$ in final form.

The significance of formula (3) and specific characteristics of its application can best be shown by definite examples. This formula is known as the formula of *integration by parts* *).

Example 1. We have already said that $\int \ln x \, dx$ is not known to us, for we do not know a function whose derivative is equal to $\ln x$. By using the formula for integration by parts we can readily find this primitive. In order to use the formula (3) the integrand must be represented in the form of a product of two functions, the first of which stands in place of u and the second in place of v' . This can be done in an infinite number of ways and we must learn to select one such method which would represent the primitive $\int uv' \, dx$ on the righthand side of formula (3) in the simplest possible form. But how can this be foreseen? Let us recall that the derivative of $\ln x$ which is equal to $1/x$ is a much simpler function than the function $\ln x$ itself and since in the transition from the primitive $\int uv' \, dx$ to the primitive $\int vu' \, dx$ the function u is replaced by its derivative u' , we can considerably simplify our problem if we assume that $u = \ln x$. But it follows from $uv' = \ln x$ and $u = \ln x$ that $v' = 1$, therefore we must choose in place of v a function whose derivative is identically equal to unity; it is, of course, simplest to assume that $v = x$; hence

$$\begin{aligned} u &= \ln x, & u' &= \frac{1}{x}, \\ v' &= 1, & v &= x, \\ uv' &= \ln x, & vu' &= 1, \end{aligned}$$

and formula (3) gives :

$$\int \ln x \, dx = x \ln x - \int 1 \cdot dx = x \ln x - x + C.$$

We can therefore see that in this case formula (3) enables us to obtain the solution in its final form. It is usually advisable to test the correctness of this solution and see that the derivative of the function obtained is, in fact, equal to $\ln x$.

*) In its general form this formula does not give a final expression for the primitive $\int uv' \, dx$ but simply gives the answer in terms of another simpler primitive; it only *partially* solves the problem of integrating the product uv' by simplifying it. Hence the term "partial integration" has been used in many European countries; the term "integration by parts" which is established in many other languages as well as also in our own, is much less expressive.

By using similar considerations and the formula for integration by parts we can also find primitive of a more general form

$$\int x^\alpha \ln x \, dx,$$

where α is an arbitrary constant; if $\alpha \neq -1$ we can assume that

$$u = \ln x, \quad u' = \frac{1}{x},$$

$$v' = x^\alpha \quad v = \frac{x^{\alpha+1}}{\alpha+1},$$

$$uv' = x^\alpha \ln x, \quad vu' = \frac{x^\alpha}{\alpha+1},$$

and formula (3) gives :

$$\int x^\alpha \ln x \, dx = \frac{x^{\alpha+1}}{\alpha+1} \left(\ln x - \frac{1}{\alpha+1} \right) + C.$$

When $\alpha = -1$, we have :

$$u = \ln x, \quad u' = \frac{1}{x},$$

$$v' = \frac{1}{x}, \quad v = \ln x,$$

and formula (3) gives :

$$\int \frac{\ln x}{x} \, dx = \ln^2 x - \int \frac{\ln x}{x} \, dx.$$

In this case the primitive on the right-hand side appears to coincide with the primitive we are trying to find. Nevertheless, the obtained relation solves our problem and we get :

$$2 \int \frac{\ln x}{x} \, dx = \ln^2 x,$$

and therefore

$$\int \frac{\ln x}{x} \, dx = \frac{1}{2} \ln^2 x + C.$$

Example 2. Let us assume that we want to find the primitive

$$\int x e^x \, dx.$$

One of the two factors of the integrand remains uncharged by differentiation while the other gives 1, *i.e.* a simpler expression than the function itself; therefore we can assume that $u = x$ and $v' = e^x$; we thus obtain :

$$\begin{aligned} u &= x, & u' &= 1, \\ v' &= e^x, & v &= e^x, \\ uv' &= xe^x, & vu' &= e^x, \end{aligned}$$

and applying formula (3) we obtain :

$$\int xe^x dx = xe^x - \int e^x dx = xe^x - e^x + C,$$

and our problem is finally solved.

Let us now assume that we want to find the primitive $\int x^2 e^x dx$; using the same arguments as above we assume :

$$\begin{aligned} u &= x^2, & u' &= 2x, \\ v' &= e^x, & v &= e^x, \\ uv' &= x^2 e^x, & vu' &= 2xe^x, \end{aligned}$$

and formula (3) gives :

$$\int x^2 e^x dx = x^2 e^x - 2 \int xe^x dx;$$

since the primitive on the right-hand side is the same as the one we found above, we also succeeded in finding the primitive $\int x^2 e^x dx$. In general, if we want to find the primitive

$$\psi_n(x) = \int x^n e^x dx,$$

where n is an arbitrary natural number, we assume that

$$\begin{aligned} u &= x^n, & u' &= nx^{n-1}, \\ v' &= e^x, & v &= e^x, \\ uv' &= x^n e^x, & vu' &= nx^{n-1} e^x, \end{aligned}$$

and formula (3) gives :

$$\psi_n(x) = x^n e^x - n \int x^{n-1} e^x dx = x^n e^x - n\psi_{n-1}(x);$$

this is a reduction formula which gives a simple expression $\psi_n(x)$ in terms of $\psi_{n-1}(x)$; since we know $\psi_0(x)$ and $\psi_1(x)$, therefore with the

help of this formula we can readily find $\psi_2(x)$, $\psi_3(x)$ in succession and, in general, $\psi_n(x)$ for every n . It can readily be seen that in this case we have for every n :

$$\psi_n(x) = \int x^n e^x dx = P_n(x) e^x + C,$$

where $P_n(x)$ is a polynomial of the n th degree.

For further exercises, cf. Problem Book by B.P. Demidovich, Section III, Nos. 123-136*).

4. Replacement of the variable. We shall now consider how the formula for differentiating composite functions can be used in integral calculus. Let $f(u)$ be a function which we can integrate and $F(u)$ be one of its primitives so that

$$F'(u) = f(u), \quad \int f(u) du = F(u) + C.$$

If we take the variable u in place of function of the new variable x , $u = \varphi(x)$, we have:

$$y = F(u) = F[\varphi(x)],$$

and according to the law for differentiating composite functions (assuming that the function $\varphi(x)$ is differentiable)

$$dy = F'[\varphi(x)] \varphi'(x) dx = f[\varphi(x)] d\varphi(x);$$

since this is a differential of the function $y = F[\varphi(x)]$, therefore conversely

$$\int f[\varphi(x)] \varphi'(x) dx = \int f[\varphi(x)] d\varphi(x) = F[\varphi(x)] + C.$$

Thus if

$$\int f(u) du = F(u) + C$$

and if $\varphi(x)$ is an arbitrary differentiable function, then

$$\int f[\varphi(x)] \varphi'(x) dx = \int f[\varphi(x)] d\varphi(x) = F[\varphi(x)] + C.$$

*) We should like to draw attention of the reader to the fact that in the Problem Book by B.P. Demidovich a primitive is called an *integral*.

In other words: if $u = \varphi(x)$ and the function $\varphi(x)$ is differentiable while the function $f(u)$ has a primitive, then

$$\int f[\varphi(x)] \varphi'(x) dx = \int f[\varphi(x)] d\varphi(x) = \int f(u) du \quad (4)$$

(where after integration it must be assumed on the right-hand side that $u = \varphi(x)$). As in integration by parts, the relation (4) only replaces finding of one primitive by another primitive; but as before, this second primitive may be simpler than the first and it may occasionally be known; in that event the first primitive can also be written down.

Every function $f(u)$ whose primitive can be found together with the relation (4) thus enables us to write an infinite number of new primitives which are obtained from the left-hand side of this relation by an arbitrary choice of the differentiable function $u = \varphi(x)$. However, because of the freedom of choosing the function $\varphi(x)$ the *method of replacing the variable* (this is the name given to the method of integration we are now considering) *) requires the development of a special inventiveness which here, as in the method of integration by parts, can only be acquired as a result of long practice. In every case the following question arises: we wish to find the primitive of a function $\varphi(x)$; in order to do this we must choose a differentiable function $\varphi(x)$ so that

$$\psi(x) = f[\varphi(x)] \varphi'(x),$$

or, which is the same,

$$\psi(x) dx = f[\varphi(x)] d\varphi(x),$$

where the primitive of the function $f(u)$

$$\int f(u) du = F(u) + C$$

is known; if we succeed in doing this, we can simply write in accordance with (4) :

$$\int \psi(x) dx = F[\varphi(x)] + C,$$

and our problem is solved. The whole difficulty of this method consists in finding the appropriate function $\varphi(x)$. In this case it is also

*) This method is also sometimes known as the substitution method.

best to illustrate our arguments by means of examples which are given below, and this might prove helpful.

Example 3. *Evaluate*

$$\int \tan x \, dx = \int \frac{\sin x \, dx}{\cos x}.$$

Owing to the fact that $\sin x \, dx = -d \cos x$, we have

$$\int \tan x \, dx = - \int \frac{d \cos x}{\cos x};$$

it is therefore natural to assume that $\cos x = u$; hence assuming that $f(u) = 1/u$, $\varphi(x) = \cos x$ in formula (4) we obtain :

$$\begin{aligned} \int \tan x \, dx &= - \int \frac{d \cos x}{\cos x} = - \int \frac{du}{u} = - \ln |u| + C = \\ &= - \ln |\cos x| + C, \end{aligned}$$

and our problem is solved.

Similarly, assuming that $u = \sin x$ we find :

$$\int \cot x \, dx = \int \frac{d \sin x}{\sin x} = \int \frac{du}{u} = \ln |u| + C = \ln |\sin x| + C.$$

Both these problems are particular cases of the following very general problem. Let $\varphi(x)$ be a differentiable function and it is required to find the primitive of the function $\varphi'(x)/\varphi(x)$. Assuming that $\varphi(x) = u$, we find that $\varphi'(x) \, dx = du$ and therefore

$$\int \frac{\varphi'(x)}{\varphi(x)} \, dx = \int \frac{du}{u} = \ln |u| + C = \ln |\varphi(x)| + C.$$

Hence

$$\int \frac{x \, dx}{1+x^2} = \frac{1}{2} \int \frac{2x \, dx}{1+x^2} = \frac{1}{2} \ln (1+x^2) + C,$$

$$\int \frac{e^x \, dx}{e^x + 1} = \ln (e^x + 1) + C,$$

and so on.

Example 4. *Evaluate the primitive*

$$\int \frac{x \, dx}{1 + \sqrt{1+x^2}}.$$

Since the numerator of the integrand is, with an accuracy upto the constant term, equal to the differential of the sum $1 + x^2$ which stands under the radical sign in the denominator, we must try to replace this radical by the new variable

$$u = \sqrt{1 + x^2};$$

hence

$$du = \frac{x dx}{\sqrt{1 + x^2}} = \frac{x dx}{u},$$

and therefore

$$x dx = u du;$$

we obtain :

$$\int \frac{x dx}{1 + \sqrt{1 + x^2}} = \int \frac{u du}{1 + u}, \quad (5)$$

i.e. we are replacing the finding of the given primitive by another much simpler primitive which, as we are now going to show, can readily be evaluated as a result of another transformation of the variable. Assuming that $1 + u = v$ we have :

$$u = v - 1, \quad du = dv,$$

and therefore

$$\begin{aligned} \int \frac{u du}{1 + u} &= \int \frac{v - 1}{v} dv = \int dv - \int \frac{dv}{v} = v - \ln v + C = \\ &= 1 + u - \ln(1 + u) + C; \end{aligned}$$

thus equation (5) gives :

$$\int \frac{x dx}{1 + \sqrt{1 + x^2}} = \sqrt{1 + x^2} - \ln(1 + x^2) + C^*,$$

where $C^* = 1 + C$ is an arbitrary constant.

We have solved our problem with the help of the transformation $u = \sqrt{1 + x^2}$, but we might have equally tried to introduce the new variable not in the form of a radical but in the form of the expression under the symbol of the radical, viz. $1 + x^2$; in this case we would have had :

$$u = 1 + x^2, \quad du = 2x dx, \quad x dx = \frac{1}{2} du,$$

and

$$\int \frac{x dx}{1 + \sqrt{1+x^2}} = \frac{1}{2} \int \frac{du}{1 + \sqrt{u}}.$$

This new primitive is also too simple. Assuming that $1 + \sqrt{u} = v$ we have :

$$u = (v - 1)^2, \quad du = 2(v - 1)dv,$$

and therefore

$$\frac{1}{2} \int \frac{du}{1 + \sqrt{u}} = \int \frac{v-1}{v} dv;$$

this is exactly the same primitive which we have obtained as a result of our first transformation of the variable. We can therefore see that in certain cases different substitutions of the variable give the same result.

Example 5. When integrating functions one frequently meets the following elementary problem : the primitive of the function $f(x)$

$$\int f(x) dx = F(x) + C$$

is known ; we are required to evaluate the primitive of the function $f(ax)$ where a is a given constant number. Let us assume that $a \neq 0$ (when $a = 0$, $f(ax) = f(0)$ is a constant and the problem becomes trivial) and that $ax = u$ so that $dx = du/a$; in this case

$$\int f(ax) dx = \int f(u) \frac{du}{a} = \frac{1}{a} F(u) + C = \frac{1}{a} F(ax) + C.$$

Therefore if

$$\int f(x) dx = F(x) + C$$

and $a \neq 0$, then

$$\int f(ax) dx = \frac{1}{a} F(ax) + C.$$

Thus

$$\begin{aligned} \int \sin kx dx &= -\frac{1}{k} \cos kx + C, \\ \int \frac{dx}{\sqrt{1-a^2x^2}} &= \frac{1}{a} \arcsin(ax) + C, \end{aligned}$$

and so on.

We are also frequently faced with the converse problem when we are unable to integrate the function $f(u)$ and we are using formula (4) in the reverse direction, as it were, by replacing the difficult expression on the right-hand side of the primitive which stands on the left-hand side of the equation; when the choice of the function $\varphi(x)$ is favourable, our problem may thus be simplified. Hence it is difficult to find the primitive

$$\int \sqrt{1 - u^2} du$$

directly; assuming that $u = \sin x$, $-\pi/2 \leq x \leq \pi/2$, we obtain, according to formula (4) :

$$\int \sqrt{1 - u^2} du = \int \cos^2 x dx ;$$

the reader can find the last primitive by himself :

$$\int \cos^2 x dx = \frac{1}{2} (x + \sin x \cos x) + C ;$$

here $x = \arcsin u$, $\sin x = u$, $\cos x = \sqrt{1 - u^2}$ and we have :

$$\int \sqrt{1 - u^2} du = \frac{1}{2} (\arcsin u + u \sqrt{1 - u^2}) + C.$$

For further exercises, cf. Problem Book by B. P. Demidovich, Section III, Nos. 28-60, 101-120.

The methods dealt in this paragraph include all the simpler methods for integration of functions; there are a few methods and, as a rule, they do not solve all the problems we are likely to encounter; these methods cannot be applied mechanically but necessitate the choice of special approaches to every problem. Nevertheless, they enable us to integrate a fairly large class of elementary functions. We shall return to this problem in chapter 16. At present we shall introduce an entirely new approach to the fundamental problem of integral calculus—a method which considerably widens and strengthens the relationship between this science and the real world—it is a mathematical apparatus for accurate study of nature and technique.

CHAPTER XII

INTEGRAL

§ 44. Area of a curvilinear trapezium

We shall now consider a number of problems encountered in different branches of science, which are inter-related by the fact that they all require a mathematical apparatus for their solution. At first this apparatus appears to have no connection with differentiation and integration of functions ; historically it developed over a long period quite independently of these two operations. However, as far back as the end of the 17th century it became clear that a general method for the solution of such problems could be developed in connection with definite problems of integral calculus. We shall soon see how this can be done.

In elementary geometry we have learnt to evaluate areas of figures bounded by straight lines and circular arcs. Areas of plane surfaces bounded by arbitrary curves can only be evaluated geometrically by means of mathematical analysis. The theoretical and practical significance of this problem is self-evident and does not require special explanations.

The figure bounded by an arbitrary curve (Fig. 27) can be divided by some mutually perpendicular lines into several parts, each of which represents a "curvilinear trapezium", *i.e.* a figure bounded on three sides by straight lines, one pair of which is parallel, and the third side is perpendicular to the other two (Fig. 28) ; the fourth side is an arc of an arbitrary curve which is intersected only at one point by an arbitrary straight line parallel to the lateral of the trapezium. We do not exclude the case (Fig. 29) when one of the two parallel sides becomes a point and we have a curvilinear triangle instead of a curvilinear trapezium. We can therefore restrict ourselves to the evaluation of area of a curvilinear trapezium.

Let us choose the system of rectangular co-ordinates such that the side of the trapezium opposite to its curvilinear side lies along the OX-axis and the trapezium itself above that axis (Fig. 30). Let us denote by a and b ($a < b$) the abscissae of the ends of the lower side ("base") of the trapezium, and let the upper curvilinear side be the function $y = f(x)$.

We are required to find the area S of our curvilinear trapezium. We must, however, remember that the area of a figure bounded by

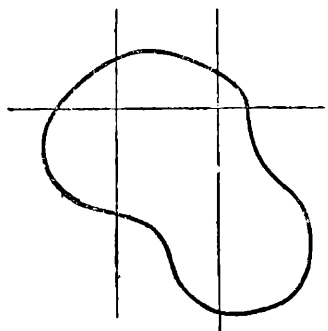


Fig. 27.

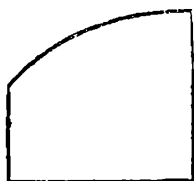


Fig. 28.

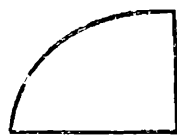


Fig. 29.

an arbitrary curve has not so far been defined — in elementary geometry this concept is only defined for figures bounded by straight lines (polygons) and for parts of circles. Hence we are faced with exactly the same type of problem as in § 26 when we were trying to define the instantaneous velocity of non-uniform motion; here, as before, we have a two-fold problem: we must define the required area and find a method for its evaluation. And again, as before, we shall solve these problems simultaneously.

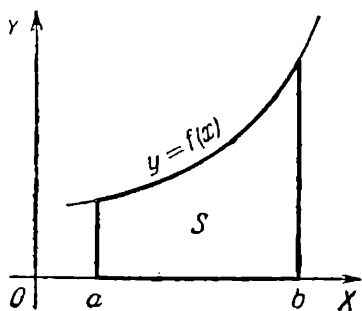


Fig. 30.

Let us recall the fact that while evaluating the area of a circle in elementary geometry we use a fundamental operation of mathematical analysis, *i.e.* the limiting process; we define the area of a circle as limit of area of a polygon which can be evaluated by elementary methods. It is therefore natural to use this method in the general case. For this purpose we divide the base (a, b) of our trapezium into an arbitrary number of parts (sections) whose points of division x_1, x_2, \dots, x_{n-1} lie between a and b , and let us assume that, in general, $a = x_0, b = x_n$ so that

$$a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b.$$

The sections (x_{k-1}, x_k) ($1 \leq k \leq n$) into which we have divided the segment (a, b) can be arbitrary in length and generally differ from one another. The set of these points of division x_k ($0 \leq k \leq n$) is called "division" of the base (a, b) .

Let us now select an arbitrary section (x_{k-1}, x_k) on the divided segment and take on it an arbitrary point ξ_k , which can either lie on this segment or coincide with one of its either ends ($x_{k-1} \leq \xi_k \leq x_k$); we draw a perpendicular from the point ξ_k to the OX-axis and produce it to intersect the curve $y = f(x)$ at the point M whose ordinate is evidently equal to $f(\xi_k)$ (Fig. 31). We draw through the point M a straight line parallel to the OX-axis to intersect the straight lines $x = x_{k-1}$ and $x = x_k$; the shaded rectangle in Fig. 31 thus has the segment $x_k - x_{k-1}$ as its base and height equal to $f(\xi_k)$, and therefore its area is equal to $f(\xi_k)(x_k - x_{k-1})$.

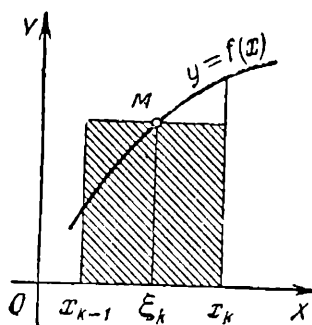


Fig. 31.

If we now repeat this construction on every segment (x_{k-1}, x_k) ($1 \leq k \leq n$), where the point ξ_k is chosen quite arbitrarily in each case on the corresponding segment, then the set of the shaded rectangles forms a "ladder"-type figure bounded by straight lines (Fig. 32). The appearance of this figure evidently depends on the division of the segment (a, b) and the chosen positions of the points ξ_k on individual sections of this division. It can readily be seen, however, that provided this division is sufficiently fine (Fig. 33), the shaded figure differs by very little from our curvilinear trapezium for all arbitrary positions of the points ξ_k . Let us denote the area of shaded figure by S^* . It is equal to sum of the areas of its component rectangles:

$$S^* = \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}).$$

If we now divide the segment (a, b) more and more finely, choose the points ξ_k arbitrarily on each section and evaluate area S^* of the shaded ladder-like figure, then it is to be expected that S^* of the shaded figure will in this process tend to a definite limit S which can be called the area of the given curvilinear trapezium. In this definition of the area S we are still influenced, in spite of our visual

representation, by the analogy with the definition of the area of a circle in elementary geometry, for in that case the area is also defined as limit of areas of polygons which approach the given circle more and more closely.

We therefore naturally arrive at the definition

$$S = \lim S^* = \lim \sum_{k=1}^n f(\xi_k) (x_k - x_{k-1}).$$

However, this is not our goal. What does the limiting process involve? How can this process be described mathematically? We know (§ 13) that a limiting process is described by the behaviour of a certain quantity which we accept as the “basic” variable. What, then, is this quantity and how does it behave in this case?

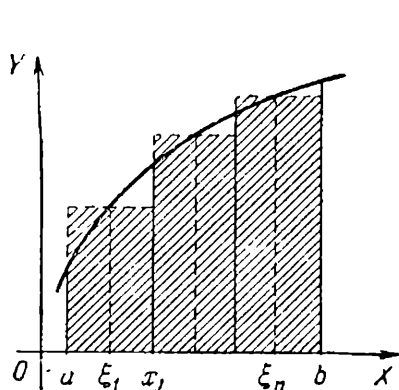


Fig. 32.

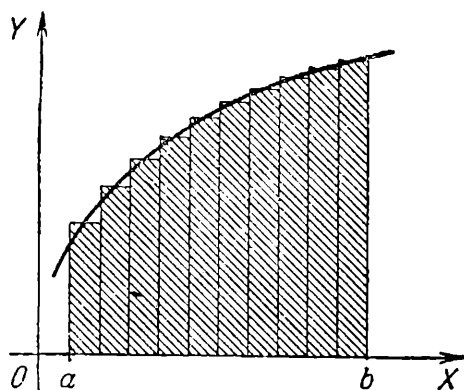


Fig. 33.

We have described the limiting process as an infinite division of the segment (a, b) . Let T be one such division; let us denote by $l(T)$ the length of the longest of the segments (x_{k-1}, x_k) ($1 \leq k \leq n$) (it is evident that for every division T this number $l(T)$ is defined uniquely). We can therefore agree that the variable division T “becomes infinitely fine” as $l(T) \rightarrow 0$, *i.e.* when the length of the longest of the segments (x_{k-1}, x_k) tends to zero, and we can therefore write

$$S = \lim_{l(T) \rightarrow 0} S^* = \lim_{l(T) \rightarrow 0} \sum_{k=1}^n f(\xi_k) (x_k - x_{k-1}). \quad (1)$$

This means that we choose $l(T)$ as the basic variable of this process and describe our process by the relation $l(T) \rightarrow 0$. It must, however, be remembered that S^* , whose limit we are trying to find, is

not a function of $l(T)$: it is obvious that one value of $l(T)$ corresponds to an infinite number of divisions T apart from the fact that even when the division T is selected, we can choose the points ξ_k in an infinite number of ways; the quantity S^* essentially depends on all these arbitrary elements and therefore it can take an infinite number of values for a given value of $l(T)$.

Hence we are here dealing with the concept of limiting process in its *wider sense* as considered in detail in § 15. The quantity S^* in which we are interested participates in a process described by the relation $l(T) \rightarrow 0$, where S^* is not a function of the basic variable $l(T)$, for it can take an infinite number of values for the given value $l(T)$. However, we know that we can nevertheless ascribe a definite limit S to the quantity S^* . The relation

$$\lim_{l(T) \rightarrow 0} S^* = S$$

has, in this case, the following exact meaning: *no matter how small $\varepsilon > 0$, we can find a $\delta > 0$ such that for every division T , where $l(T) < \delta$, and for every choice of the points ξ_k we have:*

$$|S^* - S| < \varepsilon.$$

Such a definition is quite natural. For example, if the quantity S^* tends to different limits for different divisions T or for different choices of the points ξ_k , then it would be difficult to say which of these limits measures the area of our curvilinear trapezium. It follows from our definition that the limit (1) is, in this case, absent and we can therefore ascribe no definite area to our curvilinear trapezium.

If the condition stated in our definition is satisfied, we simply say that $S^* \rightarrow S$ when the division becomes *infinitely fine*. We thus arrive at the following definition for the area of our curvilinear trapezium:

. If the sum $S^* = \sum_{k=1}^n f(\xi_k) (x_k - x_{k-1})$ tends to a definite limit

when the division becomes infinitely fine, then this limit is said to be the area of the given curvilinear trapezium.

A more formal but equivalent definition is as follows:

The number S is said to be the area of the given curvilinear trapezium

if there corresponds a $\delta > 0$ to every $\varepsilon > 0$ such that for every division T for which $l(T) < \delta$ and for every choice of the points ξ_k we have

$$\left| \sum_{k=1}^n f(\xi_k) (x_k - x_{k-1}) - S \right| < \varepsilon.$$

We have thus completely solved the first part of our problem: we have defined the area of a curvilinear trapezium. With regard to the second part, *i.e.* finding an apparatus for definite evaluation of this area, we can say that this, too, is solved in principle, for the determining formula (1) shows a succession of all operations which must be performed in order to obtain a result; however, from a practical point of view this method is not very suitable; apart from the fact that even for one definite division T and one choice of points ξ_k the evaluation of limit of such a complicated expression can only be performed in a very few simple cases; we must also remember that it is necessary to prove every time the independence of the evaluated limit from the selected system of divisions and the choice of the points ξ_k , which, in majority of cases, is very difficult. Therefore this method cannot be successfully used for the solution of specific problems. It is, however, interesting to note that in the past, when more acceptable methods of solution were unknown, these problems were solved by this direct method. Thus in ancient Greece this problem was solved when the curve $y = f(x)$ was a parabola (for example $f(x) = ax^2$, where a is a constant).

§ 45. Work done by a variable force

Let us assume that a body moves along a straight line OX under the influence of a force P which acts parallel to this line and that the direction of the force coincides with the direction in which the body moves. If this force is constant, *i.e.* its magnitude is the same at every point x on the line OX , then the work W done by this force P along any section of this straight line is, as we know, product of the force P and the length of the selected section s :

$$W = Ps.$$

Thus if the body falls from a height h on the ground under the influence of its weight P , then the force of gravity (which we can regard as constant for the duration of fall) performs the work Ph in the process.

Let us now assume that the body moves along the same line under the influence of a force which varies on different sections of the path. We can, for example, imagine that a source is situated at a point on the straight line OX which attracts or repels our body with a force P proportional to the distance between them (such are, for example, the forces of earth's gravity, the forces of electric or magnetic attraction and repulsion, *etc.*). In this case the force $P = P(x)$ is a function of the abscissa x of the point at which the body is situated at the given instant. Let us assume that the body is displaced by this variable force from the point a on the line OX to the point b on the same line. How can we find the work W done by the force P in this displacement?

We note, first of all, that we have not yet defined the concept of work due to a variable force. Hence we are again confronted with a two-fold problem—we must define this new concept in its general form and also find a practical apparatus for the evaluation of the work done by the variable force.

Let us divide the segment (a, b) arbitrarily by means of the following points of division

$$a = x_0 < x_1 < x_2 < \dots < x_n = b$$

in the same way as we did in the previous paragraph and again select an arbitrary point ξ_k on each section (x_{k-1}, x_k) ($1 \leq k \leq n$). The force acting on the body at the point ξ_k is equal to $P(\xi_k)$. If this force remains constant along the whole length of the section (x_{k-1}, x_k) , then the work along this section is equal to

$$P(\xi_k)(x_k - x_{k-1}). \quad (1)$$

In fact, however, this force varies at different points on the section (x_{k-1}, x_k) and therefore the work w_k along this section differs from that given by the above product. However, if the section (x_{k-1}, x_k) is very small, we can assume that the force $P(x)$ changes very slightly only along its length and therefore its values at different points on this section differ very little from its value at the selected point ξ_k . If this is so, we can naturally assume that the work w_k done by the force P varies very little along the section (x_{k-1}, x_k) as compared to the work done by a constant force along that section, which is equal to $P(\xi_k)$. This latter work can be expressed by the product (1), and we can therefore assume that

$$w_k \approx P(\xi_k)(x_k - x_{k-1}). \quad (2)$$

This argument can be repeated for all sections into which we have divided the segment (a, b) , *i.e.* for all k ($1 \leq k \leq n$).

Further, we are naturally also assuming that the work W done by the force P along the whole length of the segment (a, b) is equal to sum of the works along all the sections (x_{k-1}, x_k) into which we have divided the segment (a, b) , *i.e.*

$$W = \sum_{k=1}^n w_k,$$

Since we have assumed that w_k is approximately equal to the product $P(\xi_k) (x_k - x_{k-1})$, we must naturally assume further that

$$W \approx \sum_{k=1}^n P(\xi_k) (x_k - x_{k-1}). \quad (3)$$

This approximate expression for the work done will be more accurate if the approximated equations (2) are more accurate. And these equations will, in their turn, be the more accurate, the smaller the sections (x_{k-1}, x_k) are, *i.e.* the finer our division T of the segment (a, b) are. Hence we must consider the approximate equation (3) to be more accurate, the smaller the divisions T are. It will therefore be quite in order to determine the exact value of the work W done by our variable force as the limit of the sum

$$\sum_{k=1}^n P(\xi_k) (x_k - x_{k-1}),$$

as the division of the segment becomes indefinitely finer.

The structure of this sum is analogous to the sum which defines in § 44 the area of a curvilinear trapezium. And since we have again used a limiting process in order to determine the work W done by the variable force P , we can repeat word by word all arguments relating to limiting process. The formula

$$W = \lim_{l(T) \rightarrow 0} \sum_{k=1}^n P(\xi_k) (x_k - x_{k-1}) \quad (4)$$

denotes here, as before, the following : *no matter how small $\varepsilon > 0$, a $\delta > 0$ can be found such that for every division T of the segment (a, b) , provided $l(T) < \delta$, and for every arbitrary choice of the points*

$$\xi_k(x_{k-1}) \leq \xi_k \leq x_k, \quad 1 \leq k \leq n$$

we have :

$$\left| W - \sum_{k=1}^n P(\xi_k) (x_k - x_{k-1}) \right| < \varepsilon.$$

Only if this condition is satisfied, we can say that the number W gives the work done by the force P along the segment (a, b) ; however, if, for example, for different systems of division or for different selections of the points ξ_k we obtain different limits for the sum

$$\sum_{k=1}^n P(\xi_k) (x_k - x_{k-1}),$$

then it would be preferable to assume that the work done by the force P along the segment (a, b) has no definite value.

We thus see again that the first part of our problem (definition of the general concept of work done by a variable force) is completely solved whereas the second part (finding an apparatus for the practical evaluation of this work) is only solved in principle : the practical unsuitability of formula (4) justifies the repetition of all arguments used in § 44 in connection with an analogous formula.

§ 46. General concept of an integral

We have considered in §§ 44 and 45 two problems belonging to two different branches of science—one to geometry and the other to physics. While disregarding the actual contents of these problems and concentrating our attention on their analytical structure we can see that they resemble each other very closely. In either case the solution of the problem involves evaluation of limit of a sum of definite structure.

In geometry, physics, technical processes, science and in other fields of human activity many problems occur such that their analytical structure closely resembles the structure of the problems considered

above; in future we shall frequently deal with problems of this kind. It can, therefore, readily be understood that the limiting process of the described type deserves special attention and must be studied in all its aspects, for it is one of the most important problems of mathematical analysis. We shall now study it in detail.

Let the function $f(x)$ be defined in the interval (a, b) . Let us subject this interval to a division T by means of the following points of division

$$a = x_0 < x_1 < \dots < x_n = b,$$

and denote the length of the longest section (x_{k-1}, x_k) ($k = 1, 2, \dots, n$) by $l(T)$. Let us then select on every section (x_{k-1}, x_k) an arbitrary point ξ_k ($x_{k-1} \leq \xi_k \leq x_k$) and construct the sum

$$S = \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1});$$

this sum evidently depends on our division T and on the choice of the points ξ_k .

Let us now agree to say that *the sum S , as the division T becomes indefinitely finer (or, which is the same, as $l(T) \rightarrow 0$), tends to the limit I if the following condition is satisfied: no matter how small $\epsilon > 0$, a $\delta > 0$ can be found such that for every division T and provided $l(T) < \delta$ we have for every arbitrary choice of the points ξ_k*

$$|S - I| < \epsilon.$$

This fact can be denoted as follows

$$\lim_{l(T) \rightarrow 0} S = \lim_{l(T) \rightarrow 0} \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}) = I.$$

The number I , in case it exists, evidently depends only on the form of the function $f(x)$ and on the interval (a, b) . We shall call it *integral* of the function $f(x)$ from a to b [or in the interval (a, b)] and we denote it as follows:

$$\int_a^b f(x) dx; \tag{1}$$

neither the term “integral” nor the familiar symbol \int should, for the time being, be subjectively connected with the term “integration” and with the context in which this symbol was used before; we have established this connection at the same time. We should regard the notation (1) (and this point of view is justified historically) as an abbreviated notation of the sum S ; if we simply want to describe the structure of this sum without going into details as to the system of division or the choice of the point ξ_k and if we only want to fix the function f and the interval (a, b) , then we can write (disregarding accuracy of our symbolic notation)

$$S = \sum_a^b f(x) \Delta x,$$

where Δx denotes the increment of x in the transition from one point of division to the next. If we further note that $\Delta x = dx$ and denote the summation not by the Greek symbol Σ but (as it was, in fact, done in the past) by the Latin letter **S**, we obtain :

$$\mathbf{S} = \mathbf{S} \int_a^b f(x) dx,$$

and we can agree to denote the limit of this expression by the same somewhat deformed summation sign S ; one such deformation can be the symbol \int and such is, in fact, its historical origin. We thus arrive at the formula

$$I = \lim \mathbf{S} = \int_a^b f(x) dx.$$

The numbers a and b are known as *limits of integration* (a being lower and b upper) and the interval (a, b) as *interval of integration*; the function $f(x)$ is known as *integrand* and the product $f(x) dx$ as *integrand expression*.

The function $f(x)$ which has an integral in the interval (a, b) is said to be *integrable* in that interval. It can readily be seen that only a function which is bound in the interval (a, b) can be integrable. In fact, if $f(x)$ is not bounded in the interval (a, b) , then for every division T it will be unbounded and at least in one of the sections $(x_{k-1}, x_k) = \Delta_k$. If, for example, $f(x)$ assumes as large a value as we

please in Δ_k , then by choosing the point ξ_k in the usual way in this section we can make $f(\xi_k)$, and therefore the sum

$$\sum_{r=1}^n f(\xi_r) (x_r - x_{r-1})$$

as large as we please; therefore this sum cannot tend to a limit as $l(T) \rightarrow 0$.

Hence the provision that the function be bounded is the necessary condition for it to be integrable in the given interval. However, this condition is not sufficient. We shall learn one very convenient rule for the necessary and sufficient condition in § 48.

We must now study methods for evaluation of integrals for as wide a class of integrands as possible. We have already said in §§ 44 and 45 that the direct method for evaluation of such integrals as limit of a sum of a definite structure is not convenient practically and can only be used in a very few simple cases; generally speaking, it is too complicated to use. We must therefore try to find other more applicable and practical methods for this purpose.

§ 47. Upper and lower sums

In § 46 we have defined an integral as limit of a sum

$$\sum_{k=1}^n f(\xi_k) \Delta_k,$$

which is frequently called an *integral sum*. The value of this sum depends on the given division T of the interval (a, b) and on the choice of the points ξ_k in the sections Δ_k . For further development of integral calculus it is convenient to introduce sums of another kind which correspond to the given division T .

Let us assume that we are given a bounded function $f(x)$ in the interval (a, b) and let M and m denote respectively its upper and lower bounds in this interval. Let us divide the interval (a, b) by means of the usual division T with points of division

$$a = x_0 < x_1 < \dots < x_n = b;$$

let us denote the section (x_{k-1}, x_k) and its length by the same symbol Δ_k . Let M_k and m_k respectively denote the upper and lower bounds

the function $f(x)$ in the section Δ_k ($1 \leq k \leq n$). We can then construct the sums

$$S(T) = \sum_{k=1}^n M_k \Delta_k, \quad s(T) = \sum_{k=1}^n m_k \Delta_k.$$

It is evident that both sums are defined uniquely by the division T and are independent of further arbitrary elements. We shall call $S(T)$ the *upper* and $s(T)$ the *lower* sum corresponding to the given division T . We must now study some properties of these sums.

1°. Owing to the fact that $m_k \leq f(\xi_k) \leq M_k$ for every choice of the point ξ_k in the section Δ_k , we have

$$s(T) = \sum_{k=1}^n m_k \Delta_k \leq \sum_{k=1}^n f(\xi_k) \Delta_k \leq \sum_{k=1}^n M_k \Delta_k = S(T)$$

i.e. every integral sum which corresponds to the given division is confined between the upper and lower sums of this division.

2°. Let $\varepsilon > 0$ be as small as we please. It follows from the definition of the upper bound that we can choose a point ξ_k in every section Δ_k such that $f(\xi_k) > M_k - \varepsilon$; but in this case

$$\sum_{k=1}^n f(\xi_k) \Delta_k > \sum_{k=1}^n M_k \Delta_k - \varepsilon \sum_{k=1}^n \Delta_k = S(T) - \varepsilon (b - a).$$

Together with the inequality

$$\sum_{k=1}^n f(\xi_k) \Delta_k \leq S(T),$$

which, as a result of 1° applies to every integral sum, this shows that *the upper sum of the division T is the upper bound of all integral sums which correspond to this division. Similarly the lower sum of the division T is the lower bound of all integral sums which correspond to this division.*

3°. Let T and T' be two arbitrary divisions of the interval (a, b) . Let Δ_k, M_k and m_k have their former meanings for the division T and let Δ'_k, M'_k and m'_k denote the corresponding values for the division T' . Finally we denote by Δ_{kl} the length of

the common part of the section Δ_k and Δ'_l and by M_{kl} and m_{kl} the upper and lower bounds of the function $f(x)$ in the section Δ_{kl} (when the section Δ_k and Δ'_l have no interior points in common, we have $\Delta_{kl} = 0$; the symbols M_{kl} and m_{kl} can, in this case, have arbitrary value, for example $M_{kl} = m_{kl} = 0$).

We evidently have :

$$\sum_l \Delta_{kl} = \Delta_k, \quad \sum_k \Delta_{kl} = \Delta'_l.$$

$$m_k \leq m_{kl} \leq M_{kl} \leq M_k \quad (l = 1, 2, \dots),$$

$$m'_l \leq m_{kl} \leq M_{kl} \leq M'_l \quad (k = 1, 2, \dots),$$

therefore

$$\begin{aligned} s(T) &= \sum_k m_k \Delta_k = \sum_k \sum_l m_{kl} \Delta_{kl} \leq \sum_k \sum_l m_{kl} \Delta_{kl} \leq \\ &\leq \sum_k \sum_l M_{kl} \Delta_{kl} \leq \sum_k \sum_l M'_l \Delta_{kl} = \\ &= \sum_l M'_l \sum_k \Delta_{kl} = \sum_l M'_l \Delta'_l = S(T'). \end{aligned}$$

This shows that *the lower sum of every division T does not exceed the upper sum of any other division T' .*

4°. It follows from 3° that the set of all lower sums is bounded from above; let its upper bound be I_0 ; similarly the set of all upper sums is bounded from below; let its lower bound be I^0 . Owing to the fact that no lower sum can exceed the upper sum, it does not exceed either the lower bound I^0 of all upper sums; but if I^0 is thus not smaller than any lower sum, I^0 cannot be smaller than the upper bound I_0 of all lower sums. Hence we always have $I_0 \leq I^0$, i.e. *the upper bound of all lower sums does not exceed the lower bound of all upper sums.*

By using the established properties of upper and lower sums we can in the next paragraph deduce many important properties of functions in which we are interested.

§ 48. Integrability of functions

We shall follow the notation used in the previous two paragraphs. At first we shall prove the following necessary and sufficient condition for integrability of a function in an interval.

Theorem 1. (criterion of integrability). *In order that the bounded function $f(x)$ should be integrable in the interval (a, b) it is necessary and sufficient that*

$$\lim_{l(T) \rightarrow 0} [S(T) - s(T)] = 0.$$

Note 1. As always, the relation (1) means: no matter how small $\varepsilon > 0$, a $\delta > 0$ can be found such that for any division T of the interval (a, b) which satisfies the inequality $l(T) < \delta$ the inequality $S(T) - s(T) < \varepsilon$ also holds.

Note 2. The difference $\omega_k = M_k - m_k$ between the upper and lower bounds of the function $f(x)$ in the section Δ_k is known as its variation in that interval. It follows from the definition of the sums $S(T)$ and $s(T)$ that the relation (1) can also be written in the form

$$\lim_{l(T) \rightarrow 0} \sum_{k=1}^n \omega_k \Delta_k = 0.$$

Proof. 1. *Necessity.* Let it be given that the function $f(x)$ is integrable in the interval (a, b) ; let us denote its integral by I . For any division T with a sufficiently fine $l(T)$ any integral sum $\Sigma(T)$ will differ from I by less than ε and it follows 2° § 47 that $S(T)$ and $s(T)$ are respectively the upper and lower bounds of the integral sums $\Sigma(T)$; therefore none of these sums differs from I by more than ε . It therefore follows that

$$S(T) - s(T) \leq 2\varepsilon,$$

the only condition being that $l(T)$ is sufficiently small, and since ε is as small as we please, the relation (1) is proved.

2. *Sufficiency.* Let the function $f(x)$ satisfy the relation (1) in the interval (a, b) . It follows from 4° § 47 that for every division T we have

$$s(T) \leq I_0 \leq I^0 \leq S(T),$$

and for a sufficiently small $l(T)$ the difference $S(T) - s(T)$ is as small as we please; therefore $I_0 = I^0$; denoting by I the common value of these two quantities we obtain for any division T

$$s(T) \leq I \leq S(T),$$

and it also follows from 1° § 47 that

$$s(T) \leq \Sigma(T) \leq S(T),$$

where $\Sigma(T)$ is any integral sum which corresponds to the division T . It follows from the last inequalities that

$$|\Sigma(T) - I| \leq S(T) - s(T),$$

where T is an arbitrary division of the interval (a, b) and $\Sigma(T)$ is an arbitrary integral sum which corresponds to this division. But provided $l(T)$ is sufficiently small, $S(T) - s(T)$ becomes as small as we please as a result of (1); the same condition therefore also applies to $|\Sigma(T) - I|$, and this means that I is the integral of the function $f(x)$ in the interval (a, b) . Hence theorem 1 is fully proved.

Theorem 2. *If the bounds function $f(x)$ is integrable in the interval (a, b) , then the function $|f(x)|$ is also integrable in that interval.*

Proof. Denote by ω_k and ω_k^* the respective variations of the functions $f(x)$ and $|f(x)|$ in the section Δ_k . It can readily be seen that if the bounds M_k and m_k of the function $f(x)$ have the same signs in the section Δ_k , then

$$\omega_k^* = \omega_k = M_k - m_k.$$

If, however, the signs of M_k and m_k are opposite, then

$$\omega_k^* < |M_k| + |m_k| = M_k - m_k = \omega_k.$$

Hence in all cases $\omega_k \leq \omega_k^*$. It therefore follows from

$$\sum_{k=1}^n \omega_k \Delta_k \rightarrow 0$$

that

$$\sum_{k=1}^n \omega_k^* \Delta_k \rightarrow 0,$$

and in view of theorem 1 this theorem is proved

The proved criterion of integrability (theorem 1) enables us to establish existence of integrals for very wide classes of functions.

Theorem 3. *Every function $f(x)$ which is continuous in the interval (a, b) is integrable in this interval.*

The great importance of this theorem is obvious. It shows, for example, that the curvilinear trapezium which is bounded from above (cf. § 44) by a continuous curve always has a definite area.

Proof. The function $f(x)$ is continuous in the interval (a, b) ; it follows from theorem 5 § 23 that it is uniformly continuous in that interval. This implies as follows: for every $\varepsilon > 0$, a $\delta > 0$ can be found such that if $x_2 - x_1 < \delta$ ($a \leq x_1 < x_2 \leq b$), then $|f(x_2) - f(x_1)| < \varepsilon$. Let T denote any division of the interval (a, b) for which $l(T) < \delta$. Since the function $f(x)$ is continuous, it assumes in every section Δ_k its minimum value $f(\xi'_k)$ and its maximum value $f(\xi''_k)$ (theorem 2 § 23); it is therefore evident that $f(\xi'_k) = m_k$ and $f(\xi''_k) = M_k$ so that

$$\sum_{k=1}^n \omega_k \Delta_k = \sum_{k=1}^n [f(\xi''_k) - f(\xi'_k)] \Delta_k;$$

but ξ'_k and ξ''_k belong to the same section Δ_k , whose length is less than δ ; therefore $f(\xi''_k) - f(\xi'_k) < \varepsilon$ and we have:

$$\sum_{k=1}^n \omega_k \Delta_k < \varepsilon \sum_{k=1}^n \Delta_k = \varepsilon (b - a),$$

the only condition being that $l(T) < \delta$. This means that our criterion is satisfied and therefore the function $f(x)$ is integrable in the interval (a, b) .

Theorem 4. *Every function $f(x)$ which is bounded in the interval (a, b) and has only a finite number of points of discontinuity on it is integrable in that interval.*

Let the points of discontinuity of the function $f(x)$ in the interval (a, b) be as follows (in increasing order): $\alpha_1, \alpha_2, \dots, \alpha_r$, and let ε be an arbitrary positive number. We denote by d_i the section

$$(\alpha_i - \varepsilon, \alpha_i + \varepsilon) \quad (1 \leq i \leq r),$$

and let ε be so small that these sections do not overlap in pairs. The function $f(x)$ is continuous in every section $(\alpha_{i-1} + \varepsilon, \alpha_i - \varepsilon)$ (cf. Fig. 34, where these sections are marked). Therefore as in the proof of the

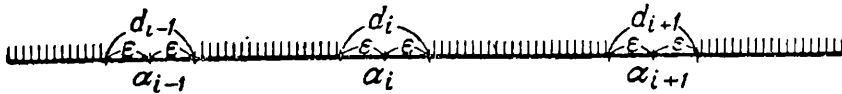


Fig. 34.

previous theorem, we can find for every section a number δ such that the variation of the function $f(x)$ in any section of length $< \delta$ which lies entirely within the marked section is smaller than ε . Evidently each marked section has its own δ ; but since there is only a finite number of such sections, each section will contain the smallest value of δ , which we shall in future denote by δ . Hence in any section of length $< \delta$ which lies entirely within one marked section (it does not matter which one) the variation of the function will be smaller than ε .

Let T denote an arbitrary division of the interval (a, b) for which $l(T) < \delta$. Let us divide the sections Δ_k of this type into two kinds:

- 1) sections of the first kind which lie entirely within one marked section, and
- 2) sections of the second kind which have points in common with another section d_i .

With this in mind let us divide the sum

$$\sum_{k=1}^n \omega_k \Delta_k = \Sigma' + \Sigma''$$

into two sums, where Σ' refers to all sections of the first kind and Σ'' to all sections of the second kind. Since $l(T) < \delta$, therefore $\omega_k < \varepsilon$ for any section of the first type and therefore

$$\Sigma' \omega_k \Delta_k < \varepsilon \Sigma' \Delta_k \leq \varepsilon \sum_{k=1}^n \Delta_k = \varepsilon (b-a). \quad (2)$$

As far as sections of the second kind are concerned, those sections which have points in common with another section d_i evidently form a section whose length is less than $2\varepsilon + 2\delta$, since they all lie in the section $(\alpha_i - \varepsilon - \delta, \alpha_i + \varepsilon + \delta)$; and since the number of sections d_i is

equal to r , the sum of the lengths of all sections of the second kind does not exceed $2r(\varepsilon + \delta)$; we can evidently always select $\delta < \varepsilon$ such that the sum of these lengths will be less than $4r\varepsilon$. And, finally, since the variation $\omega_k = M_k - m_k$ of the function $f(x)$ in any section does not exceed $M - m$ (i.e. the variation of $f(x)$ in the whole section) therefore

$$\sum'' \omega_k \Delta_k \leq (M - m) \sum'' \Delta_k \leq 4(M - m)r\varepsilon. \quad (3)$$

It follows from (2) and (3) that

$$\sum_{k=1}^n \omega_k \Delta_k < \varepsilon \{ b - a + 4(M - m)r \},$$

the only condition being that $l(T) < \delta$; since $\varepsilon > 0$ is as small as we please and the remaining letters on the right-hand side of this inequality denote constants, our criterion is satisfied and the function $f(x)$ is integrable in the interval (a, b) .

Theorem 5. *The function $f(x)$ which is bounded and monotonic in the interval (a, b) is integrable in that interval.*

This theorem does not follow directly from the preceding theorem, since a function which is bounded and monotonic in a given interval (a, b) can have an infinite number of points of discontinuity in that interval. Thus the function

$$f(x) = \begin{cases} 0 & (x = 0), \\ \frac{1}{n} & \left(\frac{1}{n+1} < x \leq \frac{1}{n} \right) (n = 1, 2, \dots), \end{cases}$$

which the student should represent graphically, is a bounded non-decreasing function in the interval $(0, 1)$; at the same time, however, it has points of discontinuity at every one of the points $1/2, 1/3, \dots, 1/n, \dots$.

Proof. Let the function $f(x)$ be non-decreasing in the interval (a, b) . It is evident that for any division T of the interval (a, b) we have the following expressions for the upper and lower bounds of the function $f(x)$ in the interval $\Delta_k = (x_{k-1}, x_k)$:

$$M_k = f(x_k), \quad m_k = f(x_{k-1}),$$

and therefore

$$\sum_{k=1}^n \omega_k \Delta_k = \sum_{k=1}^n \{ f(x_k) - f(x_{k-1}) \} \Delta_k.$$

If $l(T) < \delta$, then in this sum $\Delta_k < \delta$ ($1 \leq k \leq n$); and since the function $f(x_k) \geq f(x_{k-1})$, consequently

$$\{ f(x_k) - f(x_{k-1}) \} \Delta_k \leq \{ f(x_k) - f(x_{k-1}) \} \delta \quad (1 \leq k \leq n),$$

and this means that

$$\sum_{k=1}^n \omega_k \Delta_k \leq \delta \sum_{k=1}^n \{ f(x_k) - f(x_{k-1}) \} = \delta \{ f(b) - f(a) \}.$$

Since δ can be chosen as small as we please,

$$\sum_{k=1}^n \omega_k \Delta_k \rightarrow 0 \quad [l(T) \rightarrow 0];$$

our criterion is thus satisfied and theorem 5 is proved.

We recommend the following useful exercises from B. P. Demidovich's Problem Book : Section IV, Nos. 57, 61, 71, 73.

CHAPTER XIII

RELATIONSHIP BETWEEN AN INTEGRAL AND A PRIMITIVE

§ 49. Simple properties of integrals

In this chapter we shall try to establish the fundamental relationship existing between two basic concepts of integral calculus—the primitive and the integral, which have so far been considered quite independent of one another. In order to do this we must at first establish some simple general properties of integrals and we shall do so in this paragraph.

Theorem 1. *If the function $f(x) = c$ is constant in the interval (a, b) , then*

$$I = \int_a^b f(x) dx = \int_a^b c dx = c(b - a).$$

In order to prove this theorem it is sufficient to note that for every division T and for every choice of the points ξ_k we have

$$\sum_{k=1}^n f(\xi_k) (x_k - x_{k-1}) = c \sum_{k=1}^n (x_k - x_{k-1}) = c(b - a),$$

and therefore also

$$I = \lim_{l(T) \rightarrow 0} \sum_{k=1}^n f(\xi_k) (x_k - x_{k-1}) = c(b - a).$$

Theorem 2. If $f(x) \leq \varphi(x)$ ($a \leq x \leq b$) and $\varphi(x)$ is integrable in the interval (a, b) , then

$$\int_a^b f(x) dx \leq \int_a^b \varphi(x) dx. \quad (1)$$

In fact, it is given in the condition of this theorem that for every division T and for every choice of the points ξ_k

$$\sum_{k=1}^n f(\xi_k) \Delta_k \leq \sum_{k=1}^n \varphi(\xi_k) \Delta_k;$$

hence the limiting process for $l(T) \rightarrow 0$ gives the inequality (1).

The following is a corollary from the theorems 1 and 2.

Corollary. If the function $f(x)$ is integrable in interval (a, b) and if at an arbitrary point x in this interval $m \leq f(x) \leq M$, where m and M are arbitrary numbers, then

$$m(b-a) \leq \int_a^b f(x) dx \leq M(b-a).$$

Theorem 3. If the functions $f_1(x)$ and $f_2(x)$ are integrable in the interval (a, b) , then the function $f_1(x) \pm f_2(x)$ is also integrable in that interval, and

$$\int_a^b [f_1(x) \pm f_2(x)] dx = \int_a^b f_1(x) dx \pm \int_a^b f_2(x) dx. \quad (2)$$

In order to prove this theorem it is sufficient to note that if we denote by Σ_1 , Σ_2 and Σ the respective sums of the form

$$\sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}),$$

which is related to the functions f_1 , f_2 and $f_1 \pm f_2$ respectively, then for every division T and for every choice of the points ξ_k we evidently have :

$$\Sigma = \Sigma_1 \pm \Sigma_2,$$

and while taking limiting process for $l(T) \rightarrow 0$ we are simultaneously proving integrability of the function $f_1 \pm f_2$ and the equation (2).

Theorem 4. *If the function $f(x)$ is integrable in the interval (a, b) and α is an arbitrary constant, then the function $\alpha f(x)$ is also integrable in the interval (a, b) and*

$$\int_a^b \alpha f(x) dx = \alpha \int_a^b f(x) dx \quad (3)$$

(“the constant factor can be taken outside the symbol of integration”).

To prove this theorem it is sufficient to note that for every division of the interval (a, b) and for every choice of the points ξ_k

$$\sum_{k=1}^n \alpha f(\xi_k)(x_k - x_{k-1}) = \alpha \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1})$$

and the usual limiting process proves integrability of both the function $\alpha f(x)$ and the relation (3).

Theorem 5. *If $a \leq a' < b' \leq b$ (i.e. if the interval (a', b') comprises of a part of the interval (a, b)), then every function which is integrable in the interval (a, b) is also integrable in the interval (a', b') .*

Proof. Let it be given that the function $f(x)$ is integrable in the interval (a, b) . It follows from the criterion proved in §48 that in this case, a $\delta > 0$ corresponds to every $\epsilon > 0$ so that we always have for $l(T) < \delta$:

$$\sum_a^b = \sum_{k=1}^n \omega_k \Delta_k < \epsilon, \quad (4)$$

where the sum corresponds to the division T of the interval (a, b) . Let T' be an arbitrary division of the interval (a', b') for which $l(T') < \delta$. Let us now divide the intervals (a, a') and (b', b) into arbitrary sections of length δ ; we then evidently obtain a division T of the interval (a, b) for which $l(T) < \delta$ and therefore it follows from (4)

$$\sum_a^b < \epsilon$$

But the sum $\sum_{a'}^{b'}$ which corresponds to the division T' comprises of a

part of the sum \sum_a^b which corresponds to the division T , where all terms of the latter sum are non-negative. Therefore

$$\sum_{a'}^{b'} < \sum_a^b < \varepsilon,$$

the only condition being that $l(T') < \delta$; in terms of our criterion this means that $f(x)$ is integrable in the interval (a', b') .

Theorem 6. *Let $a < c < b$; in that case every function $f(x)$ which is integrable in each of the subintervals (a, c) and (c, b) is also integrable in the interval (a, b) and*

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (5)$$

Proof. Let T be an arbitrary division of the interval (a, b) ; we denote by T' the division obtained from T by adding the point c as a point of division and by $\Sigma(T)$ and $\Sigma(T')$ the sums of the form

$$\sum_{k=1}^n \omega_k \Delta_k,$$

which are respectively related to the divisions T and T' . These two sums differ from one another only insofar as one term of one sum is replaced by two other terms in the transition from $\Sigma(T)$ to $\Sigma(T')$; and since all terms of both sums are infinitely small for $l(T) \rightarrow 0$, therefore

$$\Sigma(T) - \Sigma(T') \rightarrow 0 \quad [l(T) \rightarrow 0]. \quad (6)$$

But the sum $\Sigma(T')$ which we have constructed in the intervals (a, b) can evidently be broken up into two similar sums for the subintervals (a, c) and (c, b) , each of which, as a result of the assumed integrability of the function $f(x)$ in subintervals (a, c) and (c, b) tends to zero for $l(T) \rightarrow 0$; hence $\Sigma(T') \rightarrow 0$; but it then follows from (6) that $\Sigma(T) \rightarrow 0$ for $l(T) \rightarrow 0$, and this means that the function $f(x)$ is integrable in the interval (a, b) .

We must still prove the relation (5) for the integrals. Since we have proved integrability of the function $f(x)$ in the interval (a, b) ,

there are no other difficulties. In fact, we have in the interval (a, b) for $l(T) \rightarrow 0$:

$$S = \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}) \rightarrow \int_a^b f(x) dx = I \quad (7)$$

independently of the divisions T and the choice of the points ξ_k ; we can, for example, select a division T such that the point c should always be a point of division; but in that case the sum S is broken up into two sums S' and S'' of a similar form constructed in the subintervals (a, c) and (c, b) respectively; it follows from the assumptions of the theorem that these two sums tend respectively to

$$\int_a^c f(x) dx = I' \quad \text{and} \quad \int_c^b f(x) dx = I''$$

for $l(T) \rightarrow 0$ so that

$$S = S' + S'' \rightarrow I' + I'' \quad (8)$$

for $l(T) \rightarrow 0$; it follows from (7) and (8) that

$$I = I' + I'',$$

and theorem 6 is thus fully proved.

Corellary. *If $a < c_1 < c_2 < \dots < c_n < b$ and the function $f(x)$ is integrable in each subinterval (a, c_1) , (c_1, c_2) , \dots , (c_n, b) , then it is integrable in the interval (a, b) and*

$$\int_a^b f(x) dx = \int_a^{c_1} f(x) dx + \int_{c_1}^{c_2} f(x) dx + \dots + \int_{c_n}^b f(x) dx.$$

As a result of theorem 5 the condition of integrability of the function $f(x)$ in each individual subinterval can be replaced by the condition of integrability of this function in the whole interval (a, b) ; it follows from theorem 5 that $f(x)$ will then also be integrable in every individual subinterval and the corollary remains valid.

Let us now make one more remark which we shall soon find very useful. The integral

$$\int_a^b f(x) dx \quad (9)$$

(if it exists) depends, according to its definition, only on the following elements :

1) the form of the function $f(x)$, and

2) the numbers a and b ;

if these elements are given, then the integral is uniquely defined. Thus, for example, the integral (9) *does not depend on the variable x* which is usually known as the “variable of integration”. Therefore by changing the symbol of this variable we do not alter the integral ; in other words, the expressions

$$\int_a^b f(x) dx, \quad \int_a^b f(y) dy, \quad \int_a^b f(\lambda) d\lambda,$$

etc. always denote one and the same thing. This simple and self-evident fact is analogous to fact that, *e.g.*

$$\sum_{k=1}^{20} \frac{1}{k}, \quad \sum_{l=1}^{20} \frac{1}{l}, \quad \sum_{\beta=1}^{20} \frac{1}{\beta},$$

etc. all denote the same thing, *viz.* the sum

$$1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{20}.$$

This sum is independent of our notation, *i.e.* of the “index of summation”, in the same way as in our example the value of the integral is independent of the symbol used for denoting the variable of integration.

§ 50. Relationship between an integral and a primitive

We shall now try to establish a law which is usually regarded as fundamental in differential and integral calculus, since it serves logically and historically as the basis for further development of these branches of mathematical analysis.

Let it be given that the function $f(x)$ is integrable in the interval (a, b) and let $a < x \leq b$; it follows from theorem 5 § 49 that the function $f(x)$ is also integrable in the subinterval (a, x) ; however, we find it inconvenient to denote its integral in that subinterval by

$$\int_a^x f(x) dx,$$

for in that case the letter x would be used in two totally different meanings : it would be used as the variable of integration and as the upper limit of the integral. Therefore using the final remark of § 49 we shall always denote the variable of integration in such cases by another letter, *i.e.*, for example, we shall write the integral of the function $f(x)$ in the subinterval (a, x) in the form

$$\int_a^x f(u) du.$$

If we now assume that the lower limit a of the integral is constant while the upper limit x can change arbitrarily in the interval (a, b) , then the above integral will evidently be a function of x which we denote by $F(x)$. We shall now prove the following fundamental proposition.

Theorem 1. *If the function $f(x)$ is integrable in the interval (a, b) and continuous at an interior point x in this interval, then the function*

$$F(x) = \int_a^x f(u) du$$

is differentiable at the point x and $F'(x) = f(x)$.

Proof. Let $a \leq x < b$ and $\Delta x > 0$ be so small that $x + \Delta x \leq b$. It then follows from theorem 6 § 49 that

$$\begin{aligned} F(x + \Delta x) - F(x) &= \int_a^{x + \Delta x} f(u) du - \int_a^x f(u) du = \\ &= \int_x^{x + \Delta x} f(u) dx^* \end{aligned} \quad (1)$$

*) So as not to exclude the case $x = a$ we should give a definite meaning to the expression $F(a) = \int_a^a f(x) dx$. It can readily be shown that $F(x) \rightarrow 0$ for $x \rightarrow a + 0$; we can therefore naturally assume that $F(a) = 0$ and we shall always do so in future.

Since it is given that the function $f(x)$ is continuous at the point x , therefore no matter how small $\varepsilon > 0$ we have for a sufficiently small Δx and $x \leq u \leq x + \Delta x$:

$$f(x) - \varepsilon < f(u) < f(x) + \varepsilon,$$

and it therefore follows from theorem 2 § 49 that

$$\int_x^{x+\Delta x} [f(x) - \varepsilon] du \leq \int_x^{x+\Delta x} f(u) du \leq \int_x^{x+\Delta x} [f(x) + \varepsilon] du.$$

Hence we obtain from the relation (1)

$$\frac{1}{\Delta x} \int_x^{x+\Delta x} [f(x) - \varepsilon] du \leq \frac{F(x + \Delta x) - F(x)}{\Delta x} \leq \frac{1}{\Delta x} \int_x^{x+\Delta x} [f(x) + \varepsilon] du.$$

In each of the two integrals the integrand is independent of the variable of integration u and it is therefore a constant. Applying theorem 1 § 49 to the right and left-hand sides of the inequalities obtained, we therefore have:

$$f(x) - \varepsilon \leq \frac{F(x + \Delta x) - F(x)}{\Delta x} \leq f(x) + \varepsilon,$$

and since ε is as small as we please for a sufficiently small Δx , these inequalities show that

$$\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = f(x).$$

Finally an analogous analysis of the case $\Delta x < 0$ (which the reader can undertake himself) readily shows that this relation remains valid for $a < x \leq b$ and $\Delta x \rightarrow -0$ and that the function $F(x)$ is therefore differentiable at the point x : hence

$$F'(x) = f(x),$$

and theorem 1 is proved.

Note. We have evidently proved more than what is necessary for the statement of theorem 1. Apart from the relation $F'(x) = f(x)$ for interior points x in the interval (a, b) , we have also established that (assuming that $F(a) = 0$)

$$\lim_{\Delta x \rightarrow +0} \frac{F(a + \Delta x) - F(a)}{\Delta x} = f(a), \quad \lim_{\Delta x \rightarrow -0} \frac{F(b + \Delta x) - F(b)}{\Delta x} = f(b)$$

(provided, of course, $f(x)$ is continuous at the points a and b respectively). It is evidently convenient to say that the function $F(x)$ is the primitive of the function $f(x)$ in the interval (a, b) if, together with the relation $F'(x) = f(x)$ for interior points in that interval, the two above mentioned limit relations are also satisfied at its end points. We shall consider this to be so in future.

Since the function $f(x)$, continuous in the interval (a, b) , is always integrable in that interval (theorem 3 §48), it follows directly from theorem 1 that

Theorem 2. *If the function $f(x)$ is continuous in the interval (a, b) , then the function*

$$F(x) = \int_a^x f(u) du$$

is the primitive of the function $f(x)$ in that interval.

In order to appreciate fully the significance of this proposition we note that the mere fact, established by this theorem, that a primitive exists for every continuous function is completely new to us; we have learnt in Chapter XI to find primitives of a few elementary functions; however, outside this narrow region of functions the problem of existence of primitives remains quite open.

If we can find integral of the given function $f(x)$ in an arbitrary subinterval (a, x) , then, as a result of theorem 2, we can also find one of the primitives $f(x)$; we know from the results of chapter XI that in this case we thus know the whole family of primitives $f(x)$. Hence, if we are able to find integral of the given function, we can also find all primitives. However, it is much more important to note that the results obtained will enable us to solve the converse problem: *by knowing one of the primitives of the continuous function $f(x)$ in the interval (a, b) we can find its integral in that interval.* In fact, let $\Phi(x)$ be an arbitrary primitive of the continuous function $f(x)$ in the interval (a, b) . Since

$$F(x) = \int_a^x f(u) du,$$

therefore, as a result of theorem 2, it is also a primitive of this function; it follows from the results of chapter XI that the difference $\Phi(x) - F(x)$ is equal to a constant number C so that

$$\Phi(x) = F(x) + C. \quad (2)$$

Assuming in this equation that $x = a$ and remembering that $F(a) = 0$ we have $C = \Phi(a)$. Therefore, assuming in (2) that $x = b$ we obtain :

$$F(b) = \int_a^b f(u) du = \Phi(b) - \Phi(a).$$

Theorem 3. *If $\Phi(x)$ is an arbitrary primitive of the function $f(x)$ in the interval (a, b) , then*

$$\int_a^b f(x) dx = \Phi(b) - \Phi(a). \quad (3)$$

Hence, by knowing any one primitive of the continuous function $f(x)$ in the interval (a, b) we can directly write its integral in that interval. The evaluation of integrals which has so many different applications therefore involves finding primitives of functions; this problem is of great importance in mathematical analysis and in its applications. We have already considered it in detail in Chapter XI and we shall consider it again in Chapters 16 and 17.

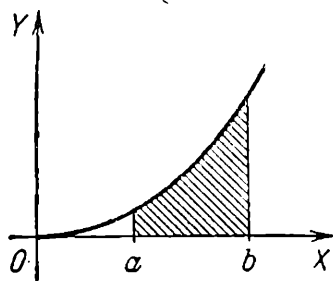


Fig. 35.

The moment when possibility of evaluating integrals by means of primitives of functions was fully realised was a turning point in the historical development of integral calculus. Prior to it the science of integration was at a very unsatisfactory level, since in each case a new method had to be found for evaluation of each individual integral; it now became possible for the first time to use a single method for many divergent classes of functions; it can therefore be said without exaggeration that from that moment onwards integral calculus began to develop as an independent scientific branch.

We shall now give one very simple method which illustrates effectiveness of formula (3). Let the curve at the top of the curvilinear trapezium (Fig. 25) be a parabola with an equation $y = cx^2 (c > 0)$

is a constant). As we have already said the area of a similar parabolic trapezium was evaluated by the ancient Greeks (Archimedes), but their method, based on the simple evaluation of limits of the corresponding sums, required complicated calculations. We are now able to write directly the complete solution of this problem by using formula (3). The required area is

$$S = \int_a^b cx^2 dx;$$

but the function cx^2 has the primitive $cx^3/3$ which we can take as the function $\Phi(x)$ in formula (3); therefore

$$S = \Phi(b) - \Phi(a) = \frac{c}{3} (b^3 - a^3),$$

which solves our problem.

For further exercises *cf.* Problem Book by B.P. Demidovich, Section IV, Nos. 1—4.

§ 51. Further properties of integrals

In § 49 we have established a series of the simple properties of integrals. We shall now add some other properties to them. In § 49 we were trying to establish properties of integrals which would lead us in the shortest possible time to formula (3) § 50, which connects the integral with the primitive of a function. Now, on the other hand, we shall use this formula to deduce some other properties of integrals.

1. In the construction of the integral

$$\int_a^b f(x) dx \quad (1)$$

we have so far always assumed that $a < b$; however, the right-hand side of formula (3) § 50 remains fully defined even when $a \geq b$. Therefore it is natural to ascribe a meaning to the expression (1) for every a and b , defining it with the help of formula (3) § 50 when $a \geq b$. Thus for every a (and for every continuous function f) we obtain, for example

$$\int_a^a f(x) dx = 0$$

and when $a > b$

$$\int_a^b f(x) dx = - \int_b^a f(x) dx \quad (2)$$

(where we also have $b > a$). The relation (2) can be stated as follows: *if the limits of integration are interchanged, the integral changes its sign.*

In all these considerations we have assumed that the function $f(x)$ is continuous in the interval (a, b) ; it is, however, natural to go further and assume that the relation (2) defines the integral $\int_a^b f(x) dx$ for $b < a$ for every function $f(x)$ which is integrable in the interval (a, b) . From now on we shall accept this definition.

Let us also draw attention to the fact that if $b = a$, the formula (3) § 50 gives:

$$\int_a^a f(x) dx = 0;$$

we have accepted this equation in § 50. We can now see that this agreement fully confirms with formula (3) § 50.

On the right-hand side of formula (3) § 50 we have the difference $\Phi(b) - \Phi(a)$ of values of the function $\Phi(x)$ for $x = b$ and $x = a$. This difference is frequently denoted as follows: $\Phi(x) \Big|_a^b$ and known as the “substitution” of the function $\Phi(x)$ from a to b .

Since, according to formula (3) § 50, the primitive can be written in the form $\int f(x) dx$, we can rewrite this formula in an equivalent form as

$$\int_a^b f(x) dx = \left(\int f(x) dx \right) \Big|_a^b. \quad (3)$$

Example 1.

$$\int_2^1 6x^2 dx = (2x^3) \Big|_2^1 = 2 - 16 = -14.$$

2. The law of integration by parts for primitives has the form

$$\int uv' dx = uv - \int vu' dx;$$

substituting on both sides from a to b we find according to the formula (3):

$$\int_a^b uv' dx = (uv) \Big|_a^b - \int_a^b vu' dx. \quad (4)$$

This is the *formula of integration by parts for integrals*.

Example 2. Let us evaluate the integral

$$\int_1^e \ln x dx,$$

and we do not remember the primitive of the function $\ln x$. Assuming

$$u = \ln x, \quad u' = \frac{1}{x},$$

$$v' = 1, \quad v = x,$$

we find with the help of formula (4):

$$\int_1^e \ln x dx = x \ln x \Big|_1^e - \int_1^e dx = e \ln e - 1 \ln 1 - (e - 1) = 1.$$

For further examples *cf.* Problem Book by B.P. Demidovich, Section IV, Nos. 15-17.

3. The method of integration by replacement of the variable is based on formula (7) § 43: when $u = \varphi(x)$, we have:

$$\int f(u) du = \int f[\varphi(x)] \varphi'(x) dx. \quad (5)$$

The left-hand side of this formula denotes the primitive $F(u)$ of the function $f(u)$ in which we have replaced u by $\varphi(x)$, *i.e.* $F[\varphi(x)]$. Therefore substituting from a to b on both sides of formula (5) we obtain:

$$\int_a^b f[\varphi(x)] \varphi'(x) dx = F[\varphi(x)] \Big|_a^b = F[\varphi(b)] - F[\varphi(a)] = \int_{\varphi(a)}^{\varphi(b)} f(u) du.$$

Hence if the function $\varphi(x)$ has a continuous derivative in the interval (a, b) and the function $f(u)$ is continuous in the interval $[\varphi(a), \varphi(b)]$, then

$$\int_a^b f[\varphi(x)] \varphi'(x) dx = \int_{\varphi(a)}^{\varphi(b)} f(u) du.$$

This is the formula of replacement of the variable for integrals.

Example 3. Assuming that $u = \varphi(x) = \cos x$ we have:

$$\int_0^{\frac{\pi}{4}} \tan x dx = - \int_0^{\frac{\pi}{4}} \frac{\varphi'(x) dx}{\varphi(x)} = - \int_{\cos 0}^{\cos \frac{\pi}{4}} \frac{du}{u} = (-\ln u) \Big|_1^{\frac{\sqrt{2}}{2}} = \frac{1}{2} \ln 2.$$

4. Mean value theorem. Let M and m denote the upper and lower bounds, respectively, of the integrable function $f(x)$ in the interval (a, b) . It follows from the corollary of theorem 2 § 49 that

$$m(b-a) \leq \int_a^b f(x) dx \leq M(b-a),$$

and therefore

$$m \leq \frac{1}{b-a} \int_a^b f(x) dx \leq M. \quad (6)$$

These inequalities apply to every function $f(x)$ integrable in the interval (a, b) . If $f(x)$ is continuous in that interval, then it follows from theorem 3 § 23 that it should assume an arbitrary value between its lowest value m and greatest value M in the interval (a, b) . But the inequality (6) shows that this condition is satisfied by the number

$$\frac{1}{b-a} \int_a^b f(x) dx;$$

therefore a point c can be found between a and b such that

$$f(c) = \frac{1}{b-a} \int_a^b f(x) dx,$$

or

$$\int_a^b f(x) dx = f(c)(b - a). \quad (7)$$

This formula does not contain anything that is essentially new: if we denote by $F(x)$ a primitive of the function $f(x)$, then formula (7) can be written in the form

$$F(b) - F(a) = F'(c)(b - a);$$

existence of the point c ($a < c < b$) which will satisfy this relation can simply be proved by applying Lagrange's theorem (§ 36) to the function $F(x)$ in the interval (a, b) . The above deduction from formula (7) is interesting insofar as the relationship between an integral and a primitive is not used.

The "mean value theorem" expressed by formula (7) can be generalised. Let the function $\varphi(x)$ be continuous and keep the same sign in the interval (a, b) ; we shall assume, say, that it is not negative. then for $a \leq x \leq b$ we have

$$m\varphi(x) \leq f(x)\varphi(x) \leq M\varphi(x),$$

and therefore it follows from theorems 4 and 2 §49 that

$$m \int_a^b \varphi(x) dx \leq \int_a^b f(x)\varphi(x) dx \leq M \int_a^b \varphi(x) dx,$$

or

$$m \leq \frac{\int_a^b f(x)\varphi(x) dx}{\int_a^b \varphi(x) dx} \leq M.$$

As before, we can therefore conclude that a point c exists between a and b for which

$$f(c) = \frac{\int_a^b f(x)\varphi(x) dx}{\int_a^b \varphi(x) dx},$$

or

$$\int_a^b f(x) \varphi(x) dx = f(c) \int_a^b \varphi(x) dx. \quad (9)$$

Theorem (Mean value theorem). *If the functions $f(x)$ and $\varphi(x)$ are continuous in the interval (a, b) and if $\varphi(x)$ keeps the same sign, then a point c can be found between a and b for which the relation (9) holds.*

We find in practice that this theorem is not as useful as the inequalities (8) which lead us to it.

Note. In the proof of formula (9) we divide by the integral

$$\int_a^b \varphi(x) dx,$$

and therefore assume that it is not equal to zero. But if the function $\varphi(x)$ is identically zero in the interval (a, b) , then the relation (9) is trivially true (for every c). If, however, $\varphi(\alpha) > 0$ only for one value of α ($a \leq \alpha < b$), then it follows from continuity of the function $\varphi(x)$ that it is also positive in the neighbourhood $(\alpha - \epsilon, \alpha + \epsilon)$ of the point α (lemma §23). If $M > 0$ denotes the lowest value of the function $\varphi(x)$ in the subinterval $(\alpha - \epsilon, \alpha + \epsilon)$, then (corollary of theorem 2 §49)

$$\int_a^b \varphi(x) dx \geq \int_{\alpha-\epsilon}^{\alpha+\epsilon} \varphi(x) dx \geq \mu [(\alpha + \epsilon) - (\alpha - \epsilon)] = 2\mu\epsilon > 0.$$

5. In practical applications one simple corollary of theorem 2 §49 is often very useful. We evidently always have

$$-|f(x)| \leq f(x) \leq |f(x)|;$$

it therefore follows from this theorem that if the function $f(x)$ is integrable in the interval $(a, b)^*$, then

$$-\int_a^b |f(x)| dx \leq \int_a^b f(x) dx \leq \int_a^b |f(x)| dx,$$

* The integrability of $|f(x)|$ is proved in §48 (theorem 2).

or, which is the same

$$\left| \int_a^b f(x) \, dx \right| \leq \int_a^b |f(x)| \, dx,$$

The absolute value of the integral does not exceed the integral of the absolute value of the integrand.

This important inequality is analogous to the inequality in elementary algebra according to which the absolute value of a sum of several numbers never exceeds the sum of their absolute values.

CHAPTER XIV GEOMETRICAL AND MECHANICAL APPLICATIONS OF INTEGRALS

§ 52. Length of an arc of a plane curve

Apart from calculating areas of plane surfaces, calculation of the lengths of arcs of plane curves is one of the most important

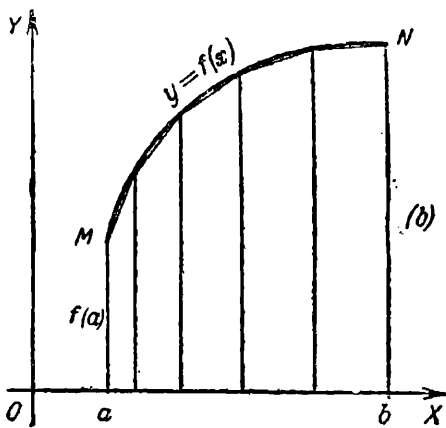


Fig. 36.

geometrical problems solved by means of integral calculus. Importance of this problem is so great from a practical point of view that no further explanations are needed. In this case, as with areas, elementary geometry enables us to calculate only the lengths of straight lines and circular arcs, but a general solution is only possible with the help of methods of mathematical analysis. In this case the logical situation again assumes its familiar

aspect : we must define simultaneously the general concept of length of an arc and find an apparatus for the practical evaluation of this length.

We shall try to solve this problem by a method which resembles even more closely than that used for calculating areas to the methods by which lengths of circumferences and arcs are calculated in elementary geometry. Let us assume that the given curve represents the graph of the function $y = f(x)$ and that we wish to find the length MN of this curve (Fig. 36) between the points $M[a, f(a)]$ and $N[b, f(b)]$. As in other problems of this type we begin with an arbitrary division T of the interval (a, b) by means of the following

points of division : $a = x_0 < x_1 < \dots < x_n = b$. From every point of division we draw a perpendicular to the OX -axis and produce it to intersect the curve $y = f(x)$. The arc MN of this curve is thus divided into n sections. Let us now connect each pair of adjacent points of division of the section MN by a rectilinear chord. The set of these chords forms a broken line inscribed in the arc MN . The length of this broken line can obviously be calculated.

If we now make the division T as fine as possible, then the constructed broken line will evidently adjoin the arc MN more and more closely. Therefore as in the case of circumference, it seems obvious to define length of the arc MN as *limit of the length of the broken line as the division T tends to become indefinitely fine*. It is, of course, essential that this limit should exist and that it should be independent of the chosen system of division T . This definition solves the first part of our problem.

The length of our constructed broken line evidently depends on the division T of the interval (a, b) ; we can therefore denote it by $L(T)$. If we denote the required length of the arc MN by L then, in accordance with our definition, we have :

$$L = \lim_{l(T) \rightarrow 0} L(T),$$

where $l(T)$ has its usual meaning (the length of the longest sub-interval of the given division). In order to obtain a method for evaluation of L on the basis of this definition it is at first necessary to find an analytical expression for the length $L(T)$ of the constructed broken line. This can readily be done. Two adjacent points of division of the arc MN have the following co-ordinates: $[x_{k-1}, f(x_{k-1})]$ and $[x_k, f(x_k)]$; therefore the length of the link in our chain connecting these two points is

$$\sqrt{(x_k - x_{k-1})^2 + [f(x_k) - f(x_{k-1})]^2},$$

and consequently

$$L(T) = \sum_{k=1} \sqrt{(x_k - x_{k-1})^2 + [f(x_k) - f(x_{k-1})]^2}.$$

Let us now assume that the function $f(x)$ has a continuous derivative $f'(x)$ in the interval (a, b) . It then follows from Lagrange's theorem that

$$f(x_k) - f(x_{k-1}) = f'(\xi_k)(x_k - x_{k-1}) \quad (1 \leq k \leq n),$$

where

$$x_{k-1} < \xi_k < x_k \quad (1 \leq k \leq n).$$

Therefore assuming further that $x_k - x_{k-1} = \Delta_k$ ($1 \leq k < n$) we have :

$$L(T) = \sum_{k=1}^n \sqrt{1 + f'^2(\xi_k)} \Delta_k.$$

Since it is given that the function $f(x)$ is continuous, therefore the function

$$\sqrt{1 + f'^2(x)} = \psi(x)$$

is also continuous and we have :

$$L(T) = \sum_{k=1}^n \psi(\xi_k) \Delta_k.$$

But we know that if $l(T) \rightarrow 0$, then irrespective of the chosen division and the position of the points ξ_k in the subintervals the following integral has as its limit

$$\int_a^b \psi(x) dx = \int_a^b \sqrt{1 + f'^2(x)} dx = \int_a^b \sqrt{1 + y'^2} dx.$$

We therefore have :

$$L = \int_a^b \sqrt{1 + y'^2} dx = \int_a^b \sqrt{1 + f'^2(x)} dx, \quad (1)$$

which fully solves our problem by replacing evaluation of the length L of the arc MN by evaluation of an integral whose integrand is known to us.

Example 1. The catenary

$$y = \cosh x = \frac{e^x + e^{-x}}{2}$$

has the form shown in Fig. 37. Let us find the length L of the arc of this curve between $x = 0$ and $x = a > 0$. We know that

$$y' = \sinh x = \frac{e^x - e^{-x}}{2},$$

and therefore

$$\sqrt{1+y'^2} = \sqrt{1+\sinh^2 x} = \cosh x$$

(this elementary calculation shows that $\cosh^2 x - \sinh^2 x = 1$ for any x). Therefore the general formula (1) gives :

$$L = \int_0^a \sqrt{1+y'^2} dx = \int_0^a \cosh x dx = \sinh x \Big|_0^a = \sinh a = \frac{e^a - e^{-a}}{2},$$

and our problem is solved.

We have so far assumed that the curve is given by an equation of the type $y = f(x)$. This means geometrically that every straight line parallel to the OY-axis intersects the section MN of the given curve at one point only. This condition may often prove to be restrictive and in some cases it may be impossible to satisfy it for any choice of coordinates — for example, when we are calculating the length of a closed curve. In such cases it is much more convenient to use the more general parametric representation of a curve by means of two equations of the type

$$x = \varphi(t), \quad y = \psi(t), \quad (2)$$

where the parameter t runs through a section $\alpha \leq t \leq \beta$ while the point (x, y) describes the section of the curve in which we are interested and which, in this case, can be of an arbitrary shape ; thus if

$$x = r \cos t, \quad y = r \sin t, \\ 0 \leq t \leq 2\pi \quad (r > 0),$$

then the point (x, y) describes a full circle

$$x^2 + y^2 = r^2$$

of radius r .

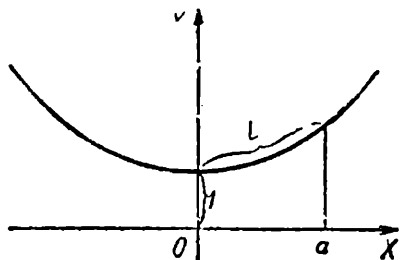


Fig. 37.

We shall now try to find an expression for the length of the arc of the given curve when this curve is expressed as parametric equations of the type (2). For this purpose we subject the interval (α, β) , along which the parameter varies, to a division T with the following points of division: $\alpha = t_0 < t_1 < \dots < t_n = \beta$. The link of the broken line which corresponds to the section $\Delta_k = t_k - t_{k-1}$ of this division evidently has the following length

$$\sqrt{[\varphi(t_k) - \varphi(t_{k-1})]^2 + [\psi(t_k) - \psi(t_{k-1})]^2}.$$

Assuming that the functions $\varphi(t)$ and $\psi(t)$ have continuous derivatives in the interval (α, β) , we have according to Lagrange's theorem :

$$\varphi(t_k) - \varphi(t_{k-1}) = \varphi'(\tau_k) \Delta_k \quad (t_{k-1} < \tau_k < t_k),$$

$$\psi(t_k) - \psi(t_{k-1}) = \psi'(\tau'_k) \Delta_k \quad (t_{k-1} < \tau'_k < t_k),$$

and therefore the length of the broken line which corresponds to the division T is

$$L(T) = \sum_{k=1}^n \sqrt{\varphi'^2(\tau_k) + \psi'^2(\tau'_k)} \Delta_k.$$

If the same value of the parameter t (for example τ_k) stands under the symbols φ'^2 and ψ'^2 , then for $l(T) \rightarrow 0$ *) the sum on the right-hand side tends, as we know, to the integral

$$L = \int_{\alpha}^{\beta} \sqrt{\varphi'^2(t) + \psi'^2(t)} dt. \quad (3)$$

In fact, however, τ'_k need not coincide with τ_k and this creates difficulty in the limiting process which we must overcome. Let us assume for the sake of brevity that

$$\sqrt{\varphi'^2(\tau_k) + \psi'^2(\tau_k)} = \rho_k, \quad \sqrt{\varphi'^2(\tau_k) + \psi'^2(\tau'_k)} = \rho'_k,$$

so that

$$L(T) = \sum_{k=1}^n \rho'_k \Delta_k = \sum_{k=1}^n \rho_k \Delta_k + \sum_{k=1}^n (\rho'_k - \rho_k) \Delta_k.$$

The first sum on the right-hand side tends to the integral (3) as its limit for $l(T) \rightarrow 0$. Hence in order to show that this integral is also limit of $L(T)$ it is sufficient to prove that the last sum on the right-hand side tends to zero for $l(T) \rightarrow 0$. We shall do so now.

If $\varphi'(\tau_k) = 0$, then $\rho_k = |\psi'(\tau_k)|$, $\rho'_k = |\psi'(\tau'_k)|$ and therefore

$$|\rho'_k - \rho_k| \leq |\psi'(\tau'_k) - \psi'(\tau_k)|$$

If, however, $\varphi'(\tau_k) \neq 0$, then $\rho_k > 0$, $\rho'_k > 0$, and

$$\rho'^2_k - \rho^2_k = \psi'^2(\tau'_k) - \psi'^2(\tau_k) = [\psi'(\tau'_k)\psi'(\tau_k)] [\psi'(\tau'_k) + \psi'(\tau_k)]$$

*) $l(T)$ denotes, as usual, the longest subinterval Δ_k of the division T .

and consequently

$$|\rho'_k - \rho_k| = \left| \frac{\psi'(\tau'_k) + \psi'(\tau_k)}{\rho'_k + \rho_k} \right| \cdot |\psi'(\tau'_k) - \psi'(\tau_k)| \leq |\psi'(\tau'_k) - \psi'(\tau_k)|,$$

so that evidently

$$\left| \frac{\psi'(\tau'_k) + \psi'(\tau_k)}{\rho'_k + \rho_k} \right| < 1.$$

We therefore have in each case :

$$|\rho'_k - \rho_k| \leq |\psi'(\tau'_k) - \psi'(\tau_k)|,$$

and therefore,

$$\left| \sum_{k=1}^n (\rho'_k - \rho_k) \Delta_k \right| \leq \sum_{k=1}^n |\psi'(\tau'_k) - \psi'(\tau_k)| \Delta_k. \quad (4)$$

But in accordance with our assumption the function $\psi'(t)$ is continuous and therefore also uniformly continuous in the interval (α, β) . If $\varepsilon > 0$ is as small as we please, then evidently

$$|\psi'(\tau'_k) - \psi'(\tau_k)| < \varepsilon \quad (1 \leq k \leq n), \quad (5)$$

provided $l(T)$ is sufficiently small. But it then follows from (4) and (5) that

$$\left| \sum_{k=1}^n (\rho'_k - \rho_k) \Delta_k \right| < \varepsilon \sum_{k=1}^n \Delta_k = \varepsilon (\beta - \alpha).$$

This proves that for $l(T) \rightarrow 0$

$$\sum_{k=1}^n (\rho'_k - \rho_k) \Delta_k \rightarrow 0,$$

and therefore

$$L(T) \rightarrow L = \int_{\alpha}^{\beta} \sqrt{\varphi'^2(t) + \psi'^2(t)} dt.$$

Hence if the curve is given in parametric form the length of the arc corresponding to the interval $\alpha \leq t \leq \beta$ of the parameter is evaluated by means of the integral (3). It is evident that this integral assumes the familiar form (1) for $t = x$.

Example 2. Find the length of the cycloid (Fig. 38)

$$x = a(t - \sin t), \quad y = a(1 - \cos t)$$

in the interval $0 \leq t \leq 2\pi$.

We have (denoting differentiation with respect to t by dashes) :

$$x' = a(1 - \cos t), \quad y' = a \sin t,$$

and consequently

$$x'^2 + y'^2 = 2a^2(1 - \cos t) = 4a^2 \sin^2 \frac{t}{2},$$

$$\sqrt{x'^2 + y'^2} = 2a \sin \frac{t}{2};$$

hence

$$L = \int_0^{2\pi} 2a \sin \frac{t}{2} dt = 2a \int_0^{\pi} 2 \sin u du = 4a(-\cos u) \Big|_0^{\pi} = 8a.$$

For further exercises *cf.* Problem Book by B. P. Demidovich, Section IV, Nos. 209, 220.

If instead of considering the interval (α, β) we consider the subinterval (α, t) , where t varies from α to β , then the length of the arc of the curve in the subinterval (α, t) will be a function of t :

$$L = L(t) = \int_{\alpha}^t \sqrt{\varphi'^2(u) + \psi'^2(u)} du$$

(as usual, in such cases we no longer denote the variable of integration by t but by any other letter, for example u).

Hence

$$L'(x) = \frac{dL}{dt} = \sqrt{\varphi'^2(t) + \psi'^2(t)} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2},$$

$$dL = \sqrt{dx^2 + dy^2}. \quad (6)$$

In the case when $t = x$, $y = f(x)$, we have similarly

$$L = L(x) = \int_a^x \sqrt{1 + f'^2(x)} dx,$$

$$L'(x) = \sqrt{1 + f'^2(x)} = \sqrt{1 + y'^2} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}.$$

Formula (6) which is independent of the choice of the parameter t shows that the differential of the length of the arc is equal to the hypotenuse of a triangle in which the sides adjacent to the right angle are equal to the differentials of the coordinates of the points of the given curve. Therefore when $t = x$, the differential of the arc of the curve in transition from the point

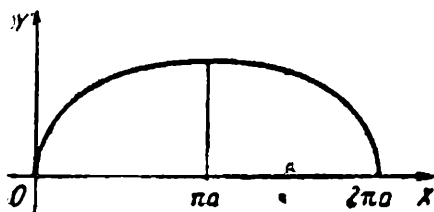


Fig. 38.

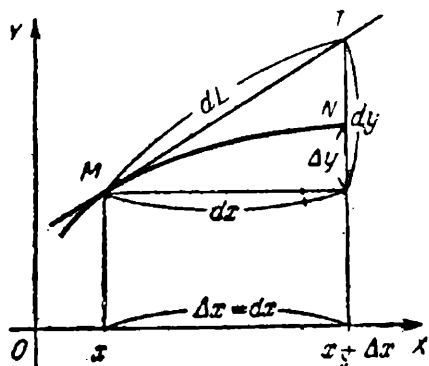


Fig. 39.

$M(x, y)$ to the point $N(x + \Delta x, y + \Delta y)$ (Fig. 39) is expressed in terms of the length MT of the section of the tangent at the point M to the given curve between the straight lines parallel to OY , which have the abscissae x and $x + \Delta x$ respectively.

Any curve which is expressed in the given interval by equations of the type (2) and has a definite length, *i.e.* for which the limit

$$L = \lim_{l(T) \rightarrow 0} L(T)$$

exists in the sense in which we have often described it, is said to be *rectified* in the given interval. It evidently follows from what is said above that every section of the circle (2) can be rectified provided the functions $\varphi(t)$ and $\psi(t)$ have continuous derivatives in this section. Such a curve can evidently be rectified also in any subinterval (α, t) ($\alpha \leq t \leq \beta$) of the interval (α, β) and its length in this subinterval is a continuous increasing function of t ; therefore, conversely, a definite value of the parameter t corresponds to every value of $L(t)$; this

follows from equation (2) and definition of a point on the given curve. In this case t , and therefore also x and y , are continuous functions of $L(t)$. For such a curve we can choose a length λ of a section of this curve as the parameter determining its points from a certain point onwards which is once and for all accepted as the origin for every other point on the curve. A definite point (x, y) on the curve corresponds to each value of λ such that the coordinates x and y of the points on the curve become continuous functions of λ :

$$x = f_1(\lambda), \quad y = f_2(\lambda); \quad (7)$$

this is evidently the parametric equation of a curve which is a particular case of the general form (2). In many cases the form (7) is particularly convenient owing to the simple geometrical meaning of parameter λ . Thus, for example, both derivatives

$$f'_1(\lambda) = \frac{dx}{d\lambda} \quad \text{and} \quad f'_2(\lambda) = \frac{dy}{d\lambda}$$

of the coordinates with respect to the parameter λ have, in this case, a simple geometrical meaning: If we assume that $L = \lambda$, then it follows from the relation (6) that:

$$\frac{dx}{d\lambda} = \frac{dx}{\sqrt{dx^2 + dy^2}} = \frac{1}{\sqrt{1 + y'^2}}, \quad \frac{dy}{d\lambda} = \frac{dy}{\sqrt{dx^2 + dy^2}} = \frac{y'}{\sqrt{1 + y'^2}},$$

where $y' = dy/dx$. If α is the angle formed by the tangent to the given curve at the given point and the positive direction of the OX -axis, then $y' = \tan \alpha$ and therefore

$$\frac{dx}{d\lambda} = \cos \alpha, \quad \frac{dy}{d\lambda} = \sin \alpha.$$

These relations can be seen directly from Fig. 39.

The established definition for rectification of a given section of the curve and also the expression (3) for the length of this section evidently depend on the choice of the parameter t in the initial equations (2). These concepts can also have a purely geometrical definition which is quite independent of the analytical representation of the given curve. On the other hand it can be shown that, provided some additional conditions are satisfied, the length of the curve (2) also becomes independent of the choice of the parameter t . However, within the scope of this course we cannot go into these details.

We have learnt above that every curve which can be expressed by an equation of the type (2) can be rectified in the given interval (α, β) , provided the functions $\varphi(t)$ and $\psi(t)$ have continuous derivatives in this interval. In practice one frequently meets the case when such an assumption cannot be made with regard to the curve but when the interval (α, β) can nevertheless be divided into a finite number of subintervals, in each of which this assumption can be made (for example the contour of a polygon). In future we shall agree to call such a curve *smooth* in the interval (α, β) .*).

It can readily be shown that *every smooth curve can be rectified*. In order to prove this, let us assume for the sake of simplicity that one "singularity" always exists, which corresponds to the value τ of the parameter t (the general case evidently creates no other difficulties), so that it is known, to begin with, that the given curve can be rectified in each of the subintervals (α, τ) and (τ, β) ; let its lengths along these sections be equal to L_1 and L_2 respectively. Let T be an arbitrary division of the interval (α, β) and T' a division resulting from the division T when the point τ is added as a point of division. In that case the broken line A' which corresponds to the division T' consists of two broken lines A'_1 and A'_2 which respectively correspond to the divisions of the subintervals (α, τ) and (τ, β) ; since the curve can be rectified in these subintervals, the lengths of the broken lines A'_1 and A'_2 are correspondingly close to L_1 and L_2 , provided the division is sufficiently fine, and therefore the length of the broken line A' is close to $L_1 + L_2$. But the length of the broken line A which corresponds to the division T differs from the length of the broken line A' only insofar as the sum of two terms of the latter, which can be as small as we please, is replaced by a single term, i.e. the two sums differ from one another by as little as we please. Hence the length of the broken line A differs by as little as we please from the length of the broken line A' which, in its turn, is very close to $L_1 + L_2$, provided the division is sufficiently fine. But this means that the given curve can be rectified in the interval (α, β) and its length is equal to $L_1 + L_2$.

*) This is evidently equivalent to the case when the curve can be expressed by equations of the type (2) in the interval (α, β) , where $\varphi(t)$ and $\psi(t)$ are continuous everywhere while $\varphi'(t)$ and $\psi'(t)$ exist and are continuous everywhere except at a finite number of points; at each of these "singularities" the function $\varphi(t)$ (as well as the $\psi(t)$) has a derivative to the right as well as to the left, but these derivatives may have different values.

Integrals can be evaluated along curves which can be rectified (and, in particular, along smooth curves) in the same way as along straight lines. Let it be given that the given rectified curve is expressed by the equations (7), where the functions f_1 and f_2 are assumed to be continuous. Let us take as origin one of the ends of the section of the given curve in which we are interested and denote by L the length of this section, so that the parameter λ varies from O to L along its length. Let us divide the interval (O, L) into subintervals by means of the following points of division

$$0 = \lambda_0 < \lambda_1 < \dots < \lambda_n = L,$$

and denote the subinterval $(\lambda_{k-1}, \lambda_k)$ ($1 \leq k \leq n$) as well as the length of this subinterval by l_k .

Let $F(x, y)$ be a function of two variables defined at all points of the given section of the curve. Select an arbitrary point λ_k^* ($\lambda_{k-1} \leq \lambda_k^* \leq \lambda_k$) and assume that $x_k = f_1(\lambda_k^*)$, $y_k = f_2(\lambda_k^*)$, so that (x_k, y_k) is an arbitrary point on that section of the given curve which corresponds to the subinterval l_k along which the parameter λ varies. Let us construct the sum

$$\sum_{k=1}^n F(x_k, y_k) l_k = \sum_{k=1}^n F[f_1(\lambda_k^*), f_2(\lambda_k^*)] l_k.$$

If the function $F[f_1(\lambda), f_2(\lambda)]$ is integrable*¹⁾ in the interval (O, L) , then the sum on the right-hand side of the last equation has the following integral as its limit, provided the division becomes indefinitely fine :

$$\int_0^L F[f_1(\lambda), f_2(\lambda)] d\lambda.$$

By denoting the whole interval (O, L) of the given curve by C , this integral can be written in the form

$$\int_C F(x, y) d\lambda$$

and is said to be *an integral of the function $F(x, y)$ along the curve C* .

*¹⁾ This will always be so if $F(x, y)$ is continuous at all points on the given curve (cf. also §88) so that $F[f_1(\lambda), f_2(\lambda)]$ is a continuous function of λ in the interval (O, L) .

Integrals taken along smooth curves occur frequently in practical applications. As simpler typical examples of this kind we shall consider in §54 problems connected with mechanical characteristics of plane curves.

§ 53. Lengths of arcs of curves in space

Evaluation of the lengths of curves in space is so similar to what has been said in the previous paragraph in connection with plane curves that we can restrict ourselves to giving the fundamental definitions and results only.

1°. If the section AB of the given curve can be expressed by the equations

$$y = f_1(x), \quad z = f_2(x) \quad (a \leq x \leq b)$$

and if the functions $f_1(x)$ and $f_2(x)$ have continuous derivatives in the interval (a, b) , then the section AB has a definite length which is equal to

$$L = \int_a^b \sqrt{1 + y'^2 + z'^2} dx = \int_a^b \sqrt{1 + f_1'^2(x) + f_2'^2(x)} dx. \quad (1)$$

2°. In general, if the section AB of the given curve can be expressed by the following parametric equations

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t) \quad (\alpha \leq t \leq \beta), \quad (2)$$

where the function $\varphi(t)$, $\psi(t)$, $\chi(t)$ have continuous derivatives in the interval (α, β) , then the section AB has a definite length which is equal to

$$L = \int_{\alpha}^{\beta} \sqrt{\varphi'^2(t) + \psi'^2(t) + \chi'^2(t)} dt. \quad (3)$$

3°. If we denote by $L(t)$ the length of the section AB of the curve in the condition 2°, which corresponds to the subinterval (α, t) of variation of the parameter $(\alpha \leq t \leq \beta)$, then

$$L'(t) = \sqrt{\varphi'^2(t) + \psi'^2(t) + \chi'^2(t)}$$

and

$$dL = \sqrt{dx^2 + dy^2 + dz^2}.$$

In particular, when we have the conditions 1°

$$L'(x) = \sqrt{1 + y'^2 + z'^2} = \sqrt{1 + f_1'^2(x) + f_2'^2(x)}.$$

4°. The curve (2) which has a definite length along the section AB can be rectified along that section. If the section AB of a continuous curve can be divided into a finite number of parts, along each of which the curve satisfies the conditions 2°, then the curve is said to be *smooth* along that section. Every smooth curve can be rectified.

5°. A curve which can be rectified can be expressed by equations of the form

$$x = \varphi(\lambda), y = \psi(\lambda), z = \chi(\lambda), \quad (4)$$

where λ is the length of the curve from a fixed origin to the point (x, y, z) . In this case

$$\frac{dx}{d\lambda} = \varphi'(\lambda) = \cos \alpha, \quad \frac{dy}{d\lambda} = \psi'(\lambda) = \cos \beta, \quad \frac{dz}{d\lambda} = \chi'(\lambda) = \cos \gamma,$$

where α, β, γ are angles between the positive direction of the axes of coordinates and the tangent to the given curve at the point (x, y, z) drawn in the direction of increasing values of λ .

6°. If a curve which can be rectified is given by an equation of the type (4) and the function $F(x, y, z)$ is continuous along the section C ($\lambda_1 \leq \lambda \leq \lambda_2$) of the curve, then the integral

$$\int_{\lambda_1}^{\lambda_2} F[\varphi(\lambda), \psi(\lambda), \chi(\lambda)] d\lambda$$

is said to be "the integral of the function $F(x, y, z)$ along the curve C " and it denotes

$$\int_C F(x, y, z) d\lambda.$$

§ 54. Mass, centre of gravity and moments of inertia of a material plane curve

1. We consider a section C of a smooth plane curve given by the following equations :

$$x = \varphi(\lambda), y = \psi(\lambda), \quad (1)$$

where λ is the length of the arc of the curve taken from the beginning of the given section so that when the point (x, y) runs along the

section C , λ increases from 0 to a number L which represents the length of the whole section under consideration. For the sake of brevity we shall in future denote the "point λ " of the given curve as the point (x, y) which corresponds to the given value of λ .

We assume that a mass is distributed along the given section of the curve ("material curve"). Let $M(\lambda)$ denote the mass distributed between the points 0 and λ on the given curve and let the function $M(\lambda)$ have a continuous derivative $M'(\lambda) = \rho(\lambda)$ in the interval $(0, L)$. While considering a *rectilinear section* in § 27, we have agreed that $\rho(\lambda)$ should be called *density* of the mass at the point λ on the given curve. The arguments which we used at the time in support of this terminology also remain valid in the general case (if, as we assume, the curve is smooth). Formerly we had to restrict ourselves to rectilinear sections only, because at that time we were not acquainted with the general concept of the length of an arc.

Hence we shall call the quantity $\rho(\lambda)$ *density* of the mass at the point λ on the given curve. Since $\rho(\lambda) = M'(\lambda)$, therefore, conversely

$$M(\lambda) = \int_0^{\lambda} \rho(u) du$$

(the lower limit of integration is so chosen that $M(0) = 0$). If we want to determine the mass

$$M(\lambda_1, \lambda_2) \quad (0 \leq \lambda_1 < \lambda_2 \leq L)$$

in the subinterval (λ_1, λ_2) of our curve, we find that

$$M(\lambda_1, \lambda_2) = M(\lambda_2) - M(\lambda_1) = \int_{\lambda_1}^{\lambda_2} \rho(u) du = \int_C \rho(x, y) d\lambda. \quad (2)$$

This formula expresses the mass of an arbitrary section of a material smooth curve which is at every point given in terms of the density $\rho(\lambda)$ of the mass.

2. If we have a system of a finite number n of material points situated in one plane, whose masses are respectively m_1, m_2, \dots, m_n and their coordinates $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, then the coordinates of the centre of gravity of this system are, as we know,

$$\bar{x} = \frac{\sum_{k=1}^n m_k x_k}{\sum_{k=1}^n m_k}, \quad \bar{y} = \frac{\sum_{k=1}^n m_k y_k}{\sum_{k=1}^n m_k}, \quad (3)$$

or, denoting by $M = \sum_{k=1}^n m_k$ the mass of the whole system,

$$\bar{x} = \frac{1}{M} \sum_{k=1}^n m_k x_k, \quad \bar{y} = \frac{1}{M} \sum_{k=1}^n m_k y_k.$$

Let us now assume that the mass is not centred at individual points but is continuously distributed along the interval $(0, L)$ of the smooth curve (1). We shall try to give a comprehensive definition of the centre of gravity of such a system and find a method for evaluating its coordinates.

Let us divide the interval $(0, L)$ into arbitrary parts ("subintervals") by means of the following points of division

$$0 = \lambda_0 < \lambda_1 < \dots < \lambda_n = L \quad (T)$$

and assume, for the sake of brevity, that $\lambda_k - \lambda_{k-1} = \Delta_k$ ($1 \leq k \leq n$). Let the density of the mass at the point λ on our curve be equal to $\rho(\lambda)$ ($0 \leq \lambda \leq L$); we assume, as before, that the function $\rho(\lambda)$ is continuous along that interval. According to formula (2) the mass of the subinterval Δ_k is equal to

$$m_k = \int_{\lambda_{k-1}}^{\lambda_k} \rho(\lambda) d\lambda,$$

and it follows from the mean value theorem (§ 51) that

$$m_k = \rho(\lambda_k^*) \Delta_k,$$

where λ_k^* is a point in the subinterval Δ_k . If the subinterval Δ_k is very small, we can imagine it to be a material point of mass m_k situated at the point λ_k^* on our curve. If we perform this replacement along every subinterval of the division (T), then our material curve will, as an approximation measure, be replaced by a system consisting of n material points; the masses and the coordinates of these points will respectively be equal to

$$m_k = \rho(\lambda_k^*) \Delta_k, \quad x_k = \varphi(\lambda_k^*), \quad y_k = \psi(\lambda_k^*) \quad (1 \leq k \leq n),$$

and it follows from formula (3) that the coordinates of the centre of gravity of this system will therefore be

$$\bar{x}(T) = \frac{\sum_{k=1}^n \rho(\lambda_k^*) \varphi(\lambda_k^*) \Delta_k}{\sum_{k=1}^n \rho(\lambda_k^*) \Delta_k}, \quad \bar{y}(T) = \frac{\sum_{k=1}^n \rho(\lambda_k^*) \psi(\lambda_k^*) \Delta_k}{\sum_{k=1}^n \rho(\lambda_k^*) \Delta_k}.$$

As the division (T) becomes indefinitely fine, the constructed fictitious system consisting of a finite number of material points resembles more and more closely our material curve. We therefore naturally assume that the coordinates (x, y) of the centre of gravity of the given material curve will be respectively equal to the limits of the number $\bar{x}(T)$ and $\bar{y}(T)$, provided the division (T) becomes indefinitely fine. This evidently gives

$$\begin{aligned} \bar{x} &= \frac{\int_0^L \rho(\lambda) \varphi(\lambda) d\lambda}{\int_0^L \rho(\lambda) d\lambda} = \frac{\int_C x \rho(x, y) d\lambda}{\int_C \rho(x, y) d\lambda}, \\ \bar{y} &= \frac{\int_0^L \rho(\lambda) \psi(\lambda) d\lambda}{\int_0^L \rho(\lambda) d\lambda} = \frac{\int_C y \rho(x, y) d\lambda}{\int_C \rho(x, y) d\lambda}, \end{aligned} \quad (4)$$

or, denoting by

$$M = \int_0^L \rho(\lambda) d\lambda$$

the mass of the whole given interval,

$$\begin{aligned} \bar{x} &= \frac{1}{M} \int_0^L \rho(\lambda) \varphi(\lambda) d\lambda = \frac{1}{M} \int_C x \rho(x, y) d\lambda, \\ \bar{y} &= \frac{1}{M} \int_0^L \rho(\lambda) \psi(\lambda) d\lambda = \frac{1}{M} \int_C y \rho(x, y) d\lambda. \end{aligned}$$

If our interval is physically *homogeneous*, i.e. if $\rho(\lambda) = \rho$ is a constant along its length, then the formulae (4) give :

$$\bar{x} = \frac{1}{L} \int_0^L \varphi(\lambda) d\lambda = \frac{1}{L} \int_C x d\lambda, \quad \bar{y} = \frac{1}{L} \int_0^L \psi(\lambda) d\lambda = \frac{1}{L} \int_C y d\lambda. \quad (5)$$

Let us again return to the mechanical system which consists of a finite number of material points. Let m_1, m_2, \dots, m_n denote the masses at these points and r_1, r_2, \dots, r_n their distances from an arbitrary axis (or from a definite point). The sum

$$K = \sum_{k=1}^n m_k r_k^2$$

is called *moment of inertia* of the given system with respect to the given axis (or point). If the points of our system are situated in one plane and have rectangular coordinates $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, then moments of inertia with respect to the OX-axis, the OY-axis and the origin O are respectively equal to

$$K_x = \sum_{k=1}^n m_k y_k^2, \quad K_y = \sum_{k=1}^n m_k x_k^2, \quad K_o = \sum_{k=1}^n m_k (x_k^2 + y_k^2).$$

Let us now assume that instead of having a system composed of a finite number of material points we have a material section of the curve (1) which we have considered above.

Maintaining all previous notations and repeating all arguments with whose help we obtained a definition for the coordinates of the centre of gravity of such a section, we readily obtain the following natural definitions for its moments of inertia with respect to the OX-axis, OY-axis and the point O:

$$\begin{aligned} K_x &= \int_0^L \rho(\lambda) \psi^2(\lambda) d\lambda = \int_C y^2 \rho(x, y) d\lambda, \\ K_y &= \int_0^L \rho(\lambda) \varphi^2(\lambda) d\lambda = \int_C x^2 \rho(x, y) d\lambda, \\ K_o &= \int_0^L \rho(\lambda) [\varphi^2(\lambda) + \psi^2(\lambda)] d\lambda = \int_C (x^2 + y^2) \rho(x, y) d\lambda. \end{aligned} \quad (6)$$

Example. Let us find the coordinates of centre of gravity of a homogeneous semicircle $x^2 + y^2 = a^2$ ($y \geq 0$) and also moments of inertia of this semicircle with respect to the diameter joining its ends. In view of symmetry it is evident that $x = 0$; therefore we have only to find \bar{y} and K_x . Denoting by λ the length of the arc of the semicircle as counted from one of its ends we can write the equations of this curve in the form

$$x = a \cos \frac{\lambda}{a}, \quad y = a \sin \frac{\lambda}{a} \quad (0 \leq \lambda \leq \pi a).$$

It therefore follows from formula (5) that

$$\bar{y} = \frac{1}{\pi a} \int_0^{\pi a} a \sin \frac{\lambda}{a} d\lambda = \frac{2}{\pi} a.$$

On the other hand, formula (6) gives:

$$K_x = \int_0^{\pi a} \rho a^2 \sin^2 \frac{\lambda}{a} d\lambda = \frac{\pi}{2} \rho a^3.$$

§ 55. Capacities of geometrical bodies

As a rule, evaluation of capacities of geometrical bodies requires more complicated analytical methods than those which we have learnt to use so far. We shall consider this problem in detail later (Chapter 27). However, many problems can be completely solved with the help of simple integration and we shall now consider such problems.

Let us assume that we want to calculate volume of the body represented in Fig 40. Let us select an arbitrary rectangular system of coordinates OXYZ in space and let us agree to call the coordinate z as "height" of the point (above the plane XOY). The plane $z = h$ (where h is a given arbitrary real number) generally intersects our body forming a definite plane figure. We shall assume that area of this "section" is known to us (or we can evaluate it in one way or the other) for every value of h .

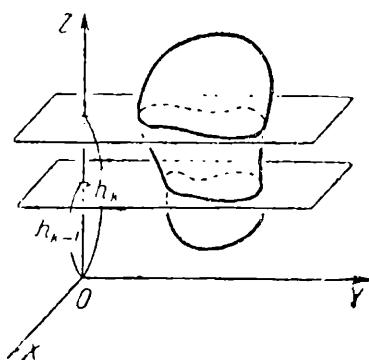


Fig. 40.

This area will in general be different for different values of h ; it is a function of h which we shall denote by $s(h)$. The special class of problems which we shall now consider is characterised by the fact that the function $s(h)$ (the area of the cross-section of the body at the height h) is assumed to be given and it is necessary to use it for expressing the volume V of the given body.

Let us at first assume that the given body is a *right cylinder*, i.e. all its horizontal sections when projected onto the XOY-plane result in the same figure (Fig. 41). If this figure is a circle, then we are dealing with a right *circular* cylinder whose volume, as we know from elementary geometry, is equal to the product of area of the base and height. We shall naturally try to adapt this rule to apply to the general case when the base of the cylinder is of an arbitrary shape *).

We therefore accept that volume of any body in the shape of a right cylinder is equal to product of the area of the base of this cylinder and its height.**))

Let us now consider the general case (Fig. 40). Let the lowest point of the body be situated at the height a and the highest point at the point b . Let us divide the interval (a, b) arbitrarily into several parts (subintervals) by means of the following points of division :

$$a = h_0 < h_1 < h_2 < \dots < h_n = b \quad (T)$$

and let us select in each subinterval

$$(h_{k-1}, h_k) = \Delta_k$$

an arbitrary point z_k such that

$$h_{k-1} \leq z_k \leq h_k \quad (1 \leq k \leq n).$$

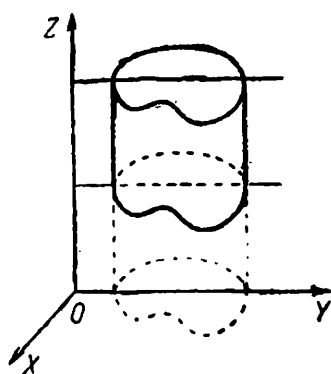


Fig. 41.

The family of planes $z = h_k$ ($k = 0, 1, \dots, n$) divides the given body into horizontal "layers" whose thicknesses are equal to $\Delta_k = h_k - h_{k-1}$ of the corresponding subinterval. If this value is very small, then the volume v_k of the k -th layer will be approximately equal to the volume of a right cylinder

*) From a logical point of view we evidently find ourselves again in the usual situation : the concept of volume of a body, except for a few cases considered in elementary geometry, has not yet been defined and it is our first duty to give a comprehensive general definition.

**) This assumption is analogous to the assumptions made earlier in connection with velocity of uniform motion, work done by a constant force, etc., when we were solving the respective problems.

of height Δ_k whose base is equal to one of the sections of the body within this layer, *e.g.* to the section of the body at the height z_k . Since area of this section is equal to $s(z_k)$, the volume of this right cylinder is equal to $s(z_k) \Delta_k$; hence we can assume approximately that

$$v_k \approx s(z_k) \Delta_k,$$

and consequently

$$V = \sum_{k=1}^n v_k \approx \sum_{k=1}^n s(z_k) \Delta_k;$$

these approximate equations will naturally be more accurate as the division (T) becomes finer; we therefore assume as usual that by definition

$$V = \lim_{l(T) \rightarrow 0} \sum_{k=1}^n s(z_k) \Delta_k,$$

provided, of course, the above limit exists and is independent of the chosen system of division and choice of the points z_k along the sections. But, as we know, this always holds, provided the function $s(h)$ is continuous in the interval (a, b) ; therefore in this case

$$V = \int_a^b s(h) dh. \quad (1)$$

This formula solves our problem and defines the general concept of volume of a body with the given cross-sectional areas; it also gives a definite method for evaluating this volume.

Example 1. Let the given body be a pyramid of height H , whose base is an arbitrary polygon of area S . We know from elementary geometry that the cross-section of this pyramid at the height h represents a polygon similar to the base whose area is proportional to the square of the distance of this section from the vertex of the pyramid, *i.e.* proportional to $(H-h)^2$.

We therefore have in this case :

$$s(h) = k (H - h)^2,$$

where k is a constant which can readily be found from the condition $s(0) = S$:

$$S = k H^2, \quad k = \frac{S}{H^2},$$

and therefore

$$s(h) = \frac{S}{H^2} (H - h)^2 = S \left(1 - \frac{h}{H}\right)^2.$$

Formula (1) gives us the following expression for volume of the pyramid

$$V = S \int_0^H \left(1 - \frac{h}{H}\right)^2 dh.$$

The substitution $1 - h/H = u$ gives

$$V = S \int_0^1 u^2 H du = \frac{SH}{3}.$$

We can thus see how easy it is to obtain with the help of integral calculus the formula which can otherwise be deduced with much greater difficulty by using the methods of elementary geometry. All arguments and results remain fully valid if the base is not a polygon but any plane figure of area S . In particular, for a cone of height H , whose base is a circle of radius R , we have the following known formula

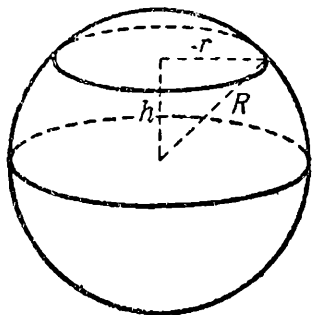


Fig. 42.

$$V = \frac{\pi R^2 H}{3}.$$

Example 2. Let the given body be a sphere of radius R , the heights h of whose cross-sectional areas are measured by their distance from the equatorial plane (Fig. 42). It can be seen from the diagram that the radius r of the section at the height h is equal to

$$r = \sqrt{R^2 - h^2},$$

therefore the cross-sectional area is

$$s(h) = \pi r^2 = \pi (R^2 - h^2),$$

and the volume of this sphere as obtained from formula (1) is

$$V = \int_{-R}^R \pi (R^2 - h^2) dh = \pi R^2 \int_{-R}^R dh - \pi \int_{-R}^R h^2 dh = \frac{4}{3} \pi R^3.$$

Example 3. Let the section $a \leq x \leq b$ of the curve $y = f(x)$ revolve about the OX-axis *) ; let us find the volume V of the body obtained as a result of this revolution (Fig. 43). All cross-sections of this body which are perpendicular to the OX-axis are evidently circles. A section by the plane $x = h$ ($a \leq h \leq b$) has, in this case, a radius equal to $f(h)$ and therefore its area is equal to $\pi [f(h)]^2$. We therefore obtain from formula (1) :

$$V = \pi \int_a^b [f(h)]^2 dh.$$

Let us also consider evaluation of the surface of the body resulting from the revolution described in example 3 ; we must remember that we have so far not defined the general concept of the surface of a curved figure, and we must therefore begin by giving this definition (at least for bodies resulting from revolution). For this purpose let us perform the usual division T of the interval (a, b) with the help of the following points of division

$$a = x_0 < x_1 < \dots < x_n = b,$$

which corresponds to the division of the given section of the curve into n parts by means of the points $[x_k, f(x_k)]$ ($k = 0, 1, \dots, n$). Join each pair of adjacent points of division $[x_{k-1}, f(x_{k-1})]$, $[x_k, f(x_k)]$ by a rectilinear chord whose length is equal to

$$l_k = \sqrt{(x_k - x_{k-1})^2 + [f(x_k) - f(x_{k-1})]^2}.$$

As the broken line formed by this chord revolves about the OX-axis, the area S^* of the surface resulting from this revolution can be regarded as approximately equal to the required area S of the surface resulting from the revolution of the given section of the curve. It is obvious that S^* will be closer to S as the division T becomes finer ; we therefore assume that

$$S = \lim_{l(T) \rightarrow 0} S^*$$

and proceed to find an analytical expression for S .

*) We assume, for the sake of simplicity, that $f(x) \geq 0$ ($a \leq x \leq b$).

We note in this connection that if the chord l_k revolves about the OX -axis, it describes a cut cone (Fig. 44) whose generating line

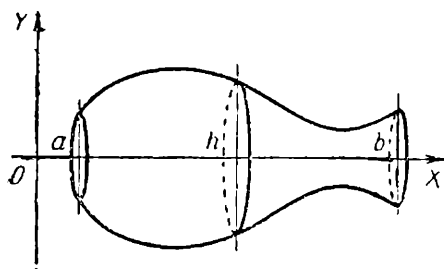


Fig. 43.

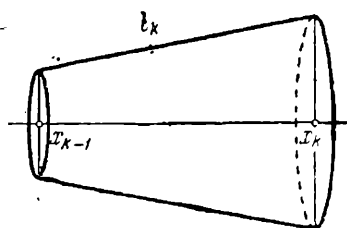


Fig. 44.

is equal to l_k and radii of the bases $f(x_{k-1})$ and $f(x_k)$ respectively. The side surface of this cut cone is equal to

$$S_k = \pi [f(x_{k-1}) + f(x_k)] l_k.$$

Let us assume that the function $f(x)$ has a continuous derivative in the interval (a, b) . It then follows from Lagrange's theorem that

$$f(x_k) - f(x_{k-1}) = f'(\xi_k) (x_k - x_{k-1}),$$

where $x_{k-1} < \xi_k < x_k$; therefore if we assume, as usual, for the sake of brevity, that $x_k - x_{k-1} = \Delta_k$ ($1 \leq k \leq n$), we obtain :

$$l_k = \sqrt{1 + f'^2(\xi_k)} \Delta_k,$$

and therefore

$$S_k = \pi [f(x_{k-1}) + f(x_k)] \sqrt{1 + f'^2(\xi_k)} \Delta_k;$$

hence

$$\begin{aligned} S^* &= \sum_{k=1}^n S_k = \pi \sum_{k=1}^n [f(x_{k-1}) + f(x_k)] \sqrt{1 + f'^2(\xi_k)} \Delta_k = \\ &= 2\pi \sum_{k=1}^n f(\xi_k) \sqrt{1 + f'^2(\xi_k)} \Delta_k + \\ &+ \pi \sum_{k=1}^n [f(x_{k-1}) + f(x_k) - 2f(\xi_k)] \sqrt{1 + f'^2(\xi_k)} \Delta_k. \end{aligned}$$

As $l(T) \rightarrow 0$, the first term on the right-hand side tends to the limit given below, since we have assumed that the function $f'(x)$ tends to the limit

$$2\pi \int_a^b f(x) \sqrt{1 + f'^2(x)} dx.$$

Therefore if we can prove that the second term on the right-hand side tends to zero for $l(T) \rightarrow 0$, then existence of a limit for S^* will be established and we have :

$$S = \lim_{l(T) \rightarrow 0} S^* = 2\pi \int_a^b f(x) \sqrt{1 + f'^2(x)} dx.$$

But it follows from the uniform continuity of the function $f(x)$ that no matter how small $\varepsilon > 0$ we have for every sufficiently fine division T :

$$|f(x_{k-1}) + f(x_k) - 2f(\xi_k)| < \varepsilon \quad (1 \leq k \leq n),$$

and consequently

$$\begin{aligned} \left| \sum_{k=1}^n [f(x_{k-1}) + f(x_k) - 2f(\xi_k)] \sqrt{1 + f'^2(\xi_k)} \Delta_k \right| &\leq \\ &\leq \varepsilon \sum_{k=1}^n \sqrt{1 + f'^2(\xi_k)} \Delta_k = \varepsilon \sum_{k=1}^n l_k \leq \varepsilon L, \end{aligned}$$

where L is the length of the given section. This evidently proves all that was required.

If we assume, for the sake of brevity, that $f(x) = y$, then we can write the following formula for the side surface of a body resulting from revolution as obtained above :

$$S = 2\pi \int_a^b y \sqrt{1 + y'^2} dx.$$

For exercises in connection with § 55 cf. Problem Book by B. P. Demidovich, Section IV, Nos. 183-185, 192, 195, 199, 200, 228.

CHAPTER XV

APPROXIMATE EVALUATION OF INTEGRALS

§ 56. Problematic setup

We have already met various definite problems of practical importance, whose solution involves evaluation of integrals. We must now probe further into what is meant by “finding” or “evaluating” an integral. If we are given the integrand and limits of integration, then the integral takes a definite numerical value ; it is finding of this numerical value which represents the problematic setup. When the numerical value of the integral can be expressed in terms of symbols generally used in mathematics (for example $5/7$, $\sqrt{2}$, $\pi^2/4$, $\sin(0.5)$, etc.), “finding” of the integral evidently implies its expression in terms of these symbols. However, most of the real numbers cannot be expressed in this final and simple symbolic way and we must therefore consider the possibility that our integral will be one such number. Numbers of this kind can only be approximately written, for example, in the form of decimal fractions with a definite number of accurate decimal places. Therefore finding of integrals evidently involves their approximate evaluation with a certain degree of accuracy. If, for example, we succeed in finding an instrument with whose help our integral can be represented as a decimal fraction with a preassigned number of accurate decimal places, we can say that our problem is solved in principle, for the term “evaluation” generally implies nothing else ; by the way, if the integral can be expressed “exactly” by one of the symbols mentioned above, then this form of expression only implies a certain definite instrument for the approximate evaluation of integrals ; thus if we find the given integral as equal to π^2 , this would enable us to represent (only by means of geometrical methods for the approximate evaluation of the number π) the given integral in the form of a decimal fraction with an arbitrary number of correct decimal places.

The instrument which enables us to evaluate the given integral approximately with an arbitrary degree of accuracy is derived from the definition of an integral. We have defined an integral as limit of sums of definite form in a definite process (when the interval of integration tends to become indefinitely small). By subjecting the basic interval to sufficiently small divisions and constructing the sums mentioned above for this division, we evidently obtain an approximate value of the integral with an arbitrary preassigned degree of accuracy. Hence, in general, a comprehensive method for evaluating integrals is already provided by its definition. And if, besides, we are looking for and keep looking for other methods leading to this goal, this is entirely due to the fact that the direct method cannot be practically applied to most of the cases because of its technical difficulties and complexities.

We have already said that the best method for which science is indebted for all its practical successes in this field is the method connecting the concept of integral with the concept of primitive of a function and we have seen in several examples how easy is to use this method for solving problems of integral calculus. If we can find the primitive $F(x)$ of the function $f(x)$ in interval (a, b) , then

$$\int_a^b f(x) dx = F(b) - F(a). \quad (1)$$

Hence evaluation of the integral in this case involves evaluation of two values of some known function. What is meant by the "known" function $F(x)$? In general, this can mean nothing but an instrument for finding the approximate value of this function with an arbitrary degree of accuracy. We have seen that in many cases this instrument is simple and convenient to apply and formula (1) readily solves our problem. What, then, is necessary to achieve success? We know (§ 50) that every continuous function $f(x)$ has a primitive. Hence formula (1) can, in principle, be used for evaluating an integral of any continuous function. However, the knowledge of existence of the function $F(x)$ is insufficient for this purpose: it is also necessary that this function should be known to us, *i.e.* we should be able to find its approximate value with an arbitrary degree of accuracy; moreover, from a practical point of view it is also necessary that the method available for the approximate evaluation should be simple and convenient, otherwise it cannot be used for practical calculations. This will always be the case when $F(x)$ belongs to the class of "elementary"

functions, since good methods for approximate evaluation are available for all such functions.

However, except for a minority of cases (to which, fortunately, belongs a fairly large class of functions which are very frequent in practice) the problem is more complicated. Differentiation of elementary functions always results in other elementary functions; however, the position is quite different for integration; elementary functions always have primitives (since all elementary functions are in general continuous), but these primitives will no longer be elementary functions. We can give many examples of simple elementary functions whose primitives are no longer elementary: for example the function $1/\ln x$, $1/\sqrt{1+x^2}$ and many others; wide classes of such functions which we shall consider later, have been discovered by P. L. Chebyshev. Let us assume, for example, that we want to evaluate the integral

$$\int_2^3 \frac{dx}{\ln x} = F(3) - F(2), \quad (2)$$

where $F(x)$ is primitive of the function $1/\ln x$. We must at first evaluate $F(3)$ and $F(2)$. But how can we do this if we do not know a convenient expression for the function $F(x)$ and also know in advance that $F(x)$ cannot be expressed in its final form in terms of elementary functions? Formula (2) does not obviously help us to evaluate the integral since all that we know about the function $F(x)$ is that it is primitive of the function $1/\ln x$ (which we know exists); we have at our disposal no other methods except the direct or indirect use of the definition of an integral.

It follows from what has been said above that integration of functions can serve as a new powerful means for defining and studying other non-elementary functions. In each case when the elementary function $f(x)$ has no elementary primitive, its primitive

$$\int_a^x f(u) du = F(x) \quad (3)$$

represents a new non-elementary function; at first we only have the definition (3) to help us study this function, *i.e.* we know nothing except that $F(x)$ is primitive of the function $f(x)$. Many functions which were primarily defined in this way received outstanding importance in the course of scientific development; the properties of

these functions were studied in detail and tables resembling logarithmic and trigonometric tables were constructed for many functions. This was the case with the function

$$\int_2^x \frac{du}{\ln u},$$

which we have considered above and which is usually denoted by $\text{Li}(x)$; it is called the "integral logarithm".

If we now return to the approximate evaluation of integrals, we can see that this problem cannot always be solved by means of primitives, and therefore it is most important to find other methods, which are more convenient from a practical point of view. These methods can be divided into two large groups: the methods of the first group are based on the original definition of an integral as limit of a sum and are perfected as far as possible so as to make them convenient for practical calculations; we shall consider in this chapter the simpler of these methods (which are sometimes called "mechanical quadratures"); the second group contains methods based on the approximate substitution of the integrand by another function whose primitive is elementary and at the same time close to the primitive of the given function; these methods necessitate application of a much more complicated apparatus of mathematical analysis and we shall consider them later (section IV).

§ 57. Method of trapeziums

The method for the approximate evaluation of integrals, usually known as the "method of trapeziums", is well illustrated in Fig. 45, where we have chosen a very rough division T of the interval of integration (a, b) . Keeping to our usual symbols we find that area of the shaded "leader-like" figure is evidently equal to the "lower sum":

$$s(T) = \sum_{k=1}^n m_k \Delta_k,$$

whereas the "upper sum"

$$S(T) = \sum_{k=1}^n M_k \Delta_k$$

is equal to the area of the surrounding "ladder-like" figure which is obtained from the shaded figure by adding the rectangles bounded from above by the dotted line. It is evident that for this rough division of the interval (a, b) both sums will differ noticeably from the integral

$$\int_a^b f(x) dx,$$

which represents the area of a curvilinear trapezium and which we are trying to evaluate approximately.

Let us join each pair of adjacent points of division of the portion of the curve $y = f(x)$ by a rectilinear chord and consider the area S of the figure bounded from above by the broken line composed of these chords, from sides by the straight lines $x = a$ and $x = b$ and from below by the OX-axis. We can already see that even with our rough division T the area S tends to come very close to the area of the curvilinear trapezium which we are trying to evaluate—in any case it comes much closer than either of the two "ladder-like" figures considered above. Therefore it is by far the most convenient to take S as the approximate value

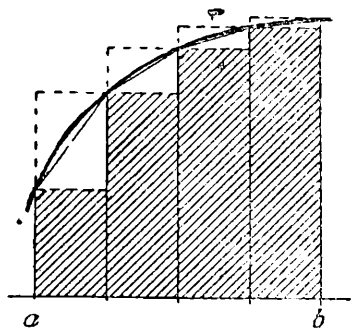


Fig. 45.

of the given integral instead of taking the upper or lower sums of the division T , particularly since, as we are going to show, the calculation of S is no more complicated than the calculation of the upper or lower sums. We must, of course, keep in mind that Fig. 45 only illustrates but does not prove anything—this is so because it shows only the positive part of the function $f(x)$ which is always convex to the same side; however, the preferential use of S instead of the upper and lower sums which it illustrates remains valid in many other cases.

As usual, let us denote by x_k and Δ_k the points and subintervals of the division T and assume for the sake of brevity that

$$f(x_k) = y_k \quad (k = 0, 1, \dots, n).$$

Let us also assume for the sake of simplicity that $f(x) \geq 0$ ($a \leq x \leq b$). The sum is sum of the areas of the rectangular trapeziums situated above the individual subintervals Δ_k (hence the name "method of trapeziums"); the trapezium, situated above the subinterval Δ_k , has

height Δ_k and bases y_{k-1} and y_k ($k = 1, 2, \dots, n$) respectively; its area is therefore equal to

$$\frac{y_{k-1} + y_k}{2} \Delta_k,$$

and therefore

$$S = \frac{1}{2} \sum_{k=1}^n (y_{k-1} + y_k) \Delta_k.$$

If we want to obtain a definite degree of accuracy, we must naturally select a sufficiently small sub-division of T (with a sufficiently small $l(T)$); otherwise the choice of points of division x_k remains arbitrary and this fact can be used to simplify the problem as far as possible. Since our formula necessitates evaluation of the function $f(x)$ at all points of division, we must at first analyse the points where the values of the function $f(x)$ will be the simplest; it may happen, for example, that this will be the case at all rational points or points which are multiples of π , etc. If there are such points, then it is evidently most convenient to select the points of division among them. If, however, the function $f(x)$ is such that it has no preferential points, then it is, of course, the simplest to divide the interval (a, b) into n equal subintervals; we thus have:

$$\Delta_k = \frac{b-a}{n} \quad (1 \leq k \leq n),$$

and we obtain

$$S = \frac{b-a}{2n} \sum_{k=1}^n (y_{k-1} + y_k) = \frac{b-a}{2} \left(\frac{y_0 + y_n}{2} + \sum_{k=1}^n y_k \right). \quad (1)$$

This is the value which we accept as the approximate value of the given integral. It can be readily seen that formula (1) remains valid in the general case when nothing is given with regard to the sign of the function $f(x)$. Naturally in order to assess the approximation given by this expression we must learn to assess the error incurred. We can see below how this is done.

Replace the curve $y = f(x)$ in an interval (α, β) ($\beta - \alpha = \Delta > 0$) by the straight chord $y = l(x)$ which joins its ends so that $l(\alpha) = f(\alpha)$, $l(\beta) = f(\beta)$. We shall try to assess the difference of the integrals

$$\int_{\alpha}^{\beta} f(x) dx - \int_{\alpha}^{\beta} l(x) dx = \int_{\alpha}^{\beta} f(x) dx - \frac{f(\alpha) + f(\beta)}{2} \Delta.$$

Let us assume for this purpose that

$$g(x) = \frac{f(x) - l(x)}{(x - \alpha)(x - \beta)} \quad (\alpha < x < \beta)$$

and consider the function

$$\varphi(z) = f(z) - l(z) - g(x)(z - \alpha)(z - \beta),$$

where x (and therefore also $g(x)$) is assumed to be constant ($\alpha < x < \beta$). We evidently have $\varphi(\alpha) = \varphi(\beta) = 0$; but it follows from the definition of $g(x)$ that $\varphi(x) = 0$, as we can readily see. Hence the function $\varphi(z)$ vanishes at the points α , β and x ($\alpha < x < \beta$); let us now assume that the function $f(x)$ has a second continuous derivative in the interval (α, β) ; the function $\varphi(z)$ will evidently also have this property. Applying Rolle's theorem to this function in the sub-intervals (α, x) and (x, β) we find that $\varphi'(z)$ vanishes twice in the interval (α, β) ; the second application of Rolle's theorem (to the function $\varphi'(z)$) shows that the function $\varphi''(z)$ also vanishes at a point ξ in the interval (α, β) . But since $l''(z) \equiv 0$, therefore

$$0 = \varphi''(\xi) = f''(\xi) - 2g(x)$$

and consequently

$$g(x) = \frac{1}{2} f''(\xi),$$

from which it also follows that $f''(\xi)$ is a continuous function of x .

We therefore find that

$$f(x) - l(x) = \frac{1}{2} f''(\xi)(x - \alpha)(x - \beta),$$

and this means that

$$\int_{\alpha}^{\beta} f(x) dx - \frac{f(\alpha) + f(\beta)}{2} \Delta = \frac{1}{2} \int_{\alpha}^{\beta} f''(\xi)(x - \alpha)(x - \beta) dx.$$

Since the function $(x - \alpha)(x - \beta)$ does not change its sign in the interval (α, β) and $f''(\xi)$ is, as we have seen above, a continuous

function of x , it follows from the mean-value theorem (§ 51) that the right hand side of the last equation can be written in the form

$$\frac{1}{2} f''(\bar{\xi}) \int_{\alpha}^{\beta} (x - \alpha)(x - \beta) dx = - \frac{(\beta - \alpha)^3}{12} f''(\bar{\xi}),$$

where $\bar{\xi}$ is an interior point of the interval (α, β) . We therefore find:

$$\int_{\alpha}^{\beta} f(x) dx - \frac{f(\alpha) + f(\beta)}{2} (\beta - \alpha) = - \frac{(\beta - \alpha)^3}{12} f''(\bar{\xi}).$$

Let us now return to the subintervals (x_{k-1}, x_k) ($1 \leq k \leq n$). If the subintervals are equal, we have

$$\alpha = x_{k-1}, \beta = x_k, f(\alpha) = y_{k-1}, f(\beta) = y_k, \Delta = \frac{b - a}{n},$$

and the formula obtained by us gives

$$\int_{x_{k-1}}^{x_k} f(x) dx - \frac{y_{k-1} + y_k}{2} (b - a) = - \frac{(b - a)^3}{12n^3} f''(\xi_k),$$

where $x_{k-1} < \xi_k < x_k$ ($1 \leq k \leq n$). Summing this equation with respect to k from 1 to n we obtain:

$$\int_a^b f(x) dx - S = - \frac{(b - a)^3}{12n^3} \sum_{k=1}^n f''(\xi_k),$$

where S is defined by formula (1).

Let m and M denote respectively the smallest and greatest values of the function $f''(x)$ in the interval (a, b) . Then we have for $1 \leq k \leq n$

$$m \leq f''(\xi_k) \leq M,$$

and also

$$nm \leq \sum_{k=1}^n f''(\xi_k) \leq nM.$$

Hence the quantity

$$\frac{1}{n} \sum_{k=1}^n f''(\xi_k)$$

is confined between m and M and therefore a point ξ^* can be found between a and b for which

$$f''(\xi^*) = \frac{1}{n} \sum_{k=1}^n f''(\xi_k),$$

and we finally obtain the following expression for the error in the formula of trapeziums

$$\int_a^b f(x) dx - \frac{b-a}{n} \left(\frac{y_0 + y_n}{2} + \sum_{k=1}^{n-1} y_k \right) = -\frac{(b-a)^3}{12n^2} f''(\xi^*).$$

We can thus see that as n increases, this error decreases in general as an infinitely small quantity of order $1/n^2$.

For exercises to § 57 cf. Problem Book by B. P. Demidovich, Section IV, Nos. 272-274.

§ 58. Method of Parabolas

The method of trapeziums is based on the fact that the curve $y=f(x)$ can be replaced (or, as is usually said, *interpolated*) by a straight chord (a linear function) in small subintervals $\Delta_k = (x_{k-1}, x_k)$. It is therefore natural to try to obtain as high an accuracy as possible by interpolating the function $y=f(x)$ in small subintervals by means of polynomials of higher degrees and preferably by means of trinomials of the second degree

$$y = \alpha x^2 + \beta x + \gamma,$$

which can usually be represented graphically as regular parabolas.

Let us divide the interval (a, b) into an *even* number $2n$ of equal parts so that

$$\Delta_k = \frac{b-a}{2n} \quad (1 \leq k \leq 2n),$$

and assume, as before, that $y_k = f(x_k)$ ($0 \leq k \leq 2n$). Let us take a pair of adjacent subintervals Δ_{2k-1} and Δ_{2k} and replace the function $y = f(x)$ so obtained in the subinterval (x_{2k-2}, x_{2k}) by the parabola

$$y = \alpha_k x^2 + \beta_k x + \gamma_k, \quad (1)$$

which passes through the points $M_{2k-2}(x_{2k-2}, y_{2k-2})$, $M_{2k-1}(x_{2k-1}, y_{2k-1})$, $M_{2k}(x_{2k}, y_{2k})$ of the given curve (Fig 46). The coefficients α_k , β_k , γ_k can evidently be evaluated from this condition; we shall not need to do so this case.

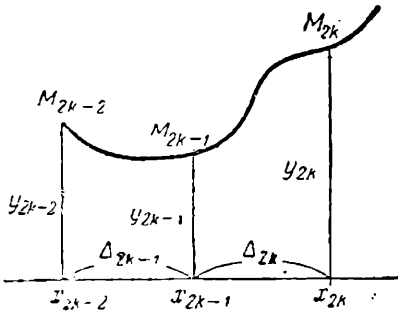


Fig 46

The method of perablas is based on the fact that in every subinterval (x_{2k-2}, x_{2k}) ($k = 1, \dots, n$) the integral of the function $f(x)$ can be approximately replaced by the integral of the correspondi parabola (1) :

$$\int_{x_{2k-2}}^{x_{2k}} f(x) dx \approx \int_{x_{2k-2}}^{x_{2k}} (\alpha_k x^2 + \beta_k x + \gamma_k) dx.$$

But in view of $x_{2k} - x_{2k-2} = (b - a) / n$ and $x_{2k} = x_{2k-2} + 2x_{2k-1}$ we obtain :

$$\begin{aligned} & \int_{x_{2k-2}}^{x_{2k}} (\alpha_k x^2 + \beta_k x + \gamma_k) dx = \\ & = \alpha_k \frac{x_{2k}^3 - x_{2k-2}^3}{3} + \beta_k \frac{x_{2k}^2 - x_{2k-2}^2}{2} + \gamma_k (x_{2k} - x_{2k-2}) = \\ & = \frac{b-a}{6n} \{ 2\alpha_k (x_{2k-2}^2 + x_{2k-2}x_{2k} + x_{2k}^2) + 3\beta_k (x_{2k-2} + x_{2k}) + 6\gamma_k \} = \\ & = \frac{b-a}{6n} \{ (\alpha_k x_{2k-2}^2 + \beta_k x_{2k-2} + \gamma_k) + (\alpha_k x_{2k}^2 + \beta_k x_{2k} + \gamma_k) + \\ & \quad + 4(\alpha_k x_{2k-1}^2 + \beta_k x_{2k-1} + \gamma_k) \} = \frac{b-a}{6n} \{ y_{2k-2} + 4y_{2k-1} + y_{2k} \}. \end{aligned}$$

Hence the method of parablas gives us :

$$\int_{x_{2k-2}}^{x_{2k}} f(x) dx \approx \frac{b-a}{6n} \{ y_{2k-2} + 4y_{2k-1} + y_{2k} \} \quad (1 \leq k \leq n),$$

and consequently summing for all subintervals

$$\begin{aligned} \int_a^b f(x) dx &\approx \frac{b-a}{6n} \sum_{k=1}^n \{y_{2k-2} + 4y_{2k-1} + y_{2k}\} = \\ &= \frac{b-a}{6n} \left\{ y_0 + y_{2n} + 4 \sum_{k=1}^n y_{2k-1} + 2 \sum_{k=1}^{n-1} y_{2k} \right\}. \end{aligned}$$

Here, as in the method of trapeziums, the assessment of error incurred in replacing the given integral by its approximate value requires special investigation which constitutes an important calculative problem. The calculation which is analogous to that performed for the method of trapeziums shows that in the case of parabolas the error generally decreases in inverse proportion to n^4 , *i.e.* much quicker than in the method of trapeziums.

For exercises to § 58 *cf.* Problem Book by B. P. Demidovich, Section IV, Nos. 275–278.

CHAPTER XVI

INTEGRATION OF RATIONAL FUNCTIONS

§ 59. Algebraic introduction

We have seen in the previous chapter that the simplest method for evaluating integrals involves finding of their primitives ; we shall now return to this problem in order to enlarge applications of this method. In view of the facts described in § 56 we shall naturally try to consider a wide class of functions whose primitives are elementary functions. At present we only know one such class of functions, *viz.* polynomials whose primitives are always polynomials ; other functions with elementary primitives which we have considered above are either isolated functions or very restricted families of functions.

The class of so-called rational functions is very closely related to the class of polynomials. These functions are said to be *rational* if their value is derived from the independent variable (and constants) by rational operations (addition, subtraction, multiplication and division) which can be repeated any number of times in any order. Hence division is only added to operations producing polynomials and this, of course, gives rise to the resulting development. We shall now learn to integrate rational functions (*i.e.* find their primitives). It is very remarkable in this connection that *primitives of all rational functions are elementary functions*. In general, these elementary functions will, of course, no longer be rational : we know that primitives of simple functions like $1/x$ and $1/(1+x^2)$ are transcendental. At the same time we shall also develop new methods for finding primitives of rational functions.

All general methodes for integrating rational functions are based on their representation in a special form convenient for integration. This representation involves algebraic operations and has no direct

connection with methods of mathematical analysis. Therefore we must begin this chapter with an algebraic introduction.

We know from elementary algebra that every rational function $f(x)$ can be represented in a definite "canonical" form

$$f(x) = \frac{P(x)}{Q(x)}, \quad (1)$$

where $P(x)$ and $Q(x)$ are polynomials with no common roots. Such fractions are usually called *rational fractions*; if the power of numerator of the fraction is lower than the power of its denominator, the fraction is said to be *regular*; otherwise it is *irregular*.

If the rational fraction (1) is irregular, then using the algebraic method for dividing polynomials we can use simple rational operations and represent our fraction in the form

$$\frac{P(x)}{Q(x)} = S(x) + \frac{R(x)}{Q(x)},$$

where $S(x)$ (quotient) and $R(x)$ (remainder) are also polynomials, but the power of remainder is always lower than the power of divisor so that the rational fraction on the right-hand side of the above equation is regular. Hence an irregular rational fraction can always be represented as sum of a polynomial and a regular fraction. And since we can integrate polynomials, integration of irregular rational fractions becomes restricted to integration of regular fractions. We can therefore only consider the case when $f(x)$ is a regular rational fraction.

In all methods for integrating rational fractions an important part is played by the roots of the denominator $Q(x)$ of this fraction. If a real or complex number α is a root of the polynomial $Q(x)$, then $Q(x)$ can be divided without remainder by the binomial $x - \alpha$, i.e.

$$Q(x) = (x - \alpha) Q^*(x),$$

where $Q^*(x)$ is also a polynomial; if $Q^*(\alpha) = 0$, we have

$$Q(x) = (x - \alpha)^2 Q^{**}(x),$$

etc. If

$$Q(x) = (x - \alpha)^k Q_1(x), \quad (2)$$

where $k \geq 1$ and $Q_1(\alpha) \neq 0$ (i.e. α is no longer a root of the polynomial $Q_1(x)$) we say that the polynomial $Q(x)$ has root α of multiplicity k .

Lemma 1. *If the real number α is a root of multiplicity $k > 0$ of the polynomial $Q(x)$, we have identically*

$$\frac{P(x)}{Q(x)} = \frac{A_k}{(x - \alpha)^k} + \frac{P_1(x)}{(x - \alpha)^{k-1} Q_1(x)}, \quad (3)$$

where A_k is a constant and $P_1(x)$ a polynomial

The polynomial $Q_1(x)$ is defined in this case by the equation (2) (so that $Q_1(\alpha) \neq 0$), the number A_k is real, all polynomials have real coefficients and the fraction on the left-hand side can be regular or irregular.

Proof. The identity (3) is equivalent to the identity

$$P(x) - A_k Q_1(x) = (x - \alpha) P_1(x), \quad (4)$$

which is obtained on multiplying by $Q(x)$; the latter identity implies that the polynomial $P(x) - A_k Q_1(x)$ can be divided by the binomial $x - \alpha$; we know that for this purpose it is necessary and sufficient that

$$P(\alpha) - A_k Q_1(\alpha) = 0. \quad (5)$$

If we therefore assume that

$$A_k = \frac{P(\alpha)}{Q_1(\alpha)}$$

(remembering that $Q_1(\alpha) \neq 0$), then the equation (5) will be satisfied and the polynomial $P(x) - A_k Q_1(x)$ will be divisible by $x - \alpha$, i.e. we shall have the identity (4) and hence also the identity (3).

If $k \geq 2$, then the rational fraction

$$\frac{P_1(x)}{(x - \alpha)^{k-1} Q_1(x)}$$

has the same form as the initial fraction $P(x) / Q(x)$; applying the proved lemma to this fraction we obtain :

$$\frac{P_1(x)}{Q_1(x)} = \frac{A_{k-1}}{(x - \alpha)^{k-1}} + \frac{P_2(x)}{(x - \alpha)^{k-2} Q_1(x)};$$

if $k \geq 3$, this process can be continued until the denominator of the last fraction on the right-hand side still contains the binomial $x - \alpha$ of an arbitrary positive power. Hence we finally obtain :

$$\frac{P(x)}{Q(x)} = \frac{A_k}{(x - \alpha)^k} + \frac{A_{k-1}}{(x - \alpha)^{k-2}} + \dots + \frac{A_1}{x - \alpha} + \frac{P^*(x)}{Q_1(x)}, \quad (6)$$

where A_1, \dots, A_k are real numbers and $P^*(x)$ is a polynomial with real coefficients.

In all these arguments we have assumed that the number α is real. Our arguments evidently remain valid for every complex number α , but the numbers A_i and the coefficients of the polynomials obtained are also complex. We did not consider and do not intend to consider integration of complex expressions; therefore when the root α is a complex number, we shall expand the given rational fraction by another method.

If the complex number $\alpha = \beta + i\gamma$ ($\gamma \neq 0$) is a root of multiplicity k of the polynomial $Q(x)$ (with real coefficients), then, as we know from algebra, the "conjugate" complex number $\alpha^* = \beta - i\gamma$ will also be a root of this polynomial of the same multiplicity k . In this case the polynomial $Q(x)$ is divisible by $(x - \alpha)^k$ and by $(x - \alpha^*)^k$ and hence also by their product; and since

$$(x - \alpha)(x - \alpha^*) = (x - \beta)^2 + \gamma^2,$$

therefore we obtain :

$$Q(x) = [(x - \beta)^2 + \gamma^2]^k Q_1(x), \quad (7)$$

where $Q_1(\alpha) \neq 0$ and $Q_1(\alpha^*) \neq 0$; the numbers β and γ and the coefficients of the polynomial $Q_1(x)$ are evidently real.

Lemma 2. *If the complex number $\alpha = \beta + i\gamma$ ($\gamma \neq 0$) is a root of multiplicity k of the polynomial $Q(x)$, then identically*

$$\frac{P(x)}{Q(x)} = \frac{B_k x + C_k}{[(x - \beta)^2 + \gamma^2]^k} + \frac{P_1(x)}{[(x - \beta)^2 + \gamma^2]^{k-1} Q_1(x)}, \quad (8)$$

where B_k and C_k are constants and $P_1(x)$ is a polynomial.

The polynomial $Q_1(x)$ is here defined by the equation (7), the numbers B_k, C_k and the coefficients of the polynomial $P_1(x)$ are real and the fraction on the left-hand side can be regular or irregular.

Proof. For the sake of brevity let us assume that

$$(x - \alpha)(x - \alpha^*) = (x - \beta)^2 + \gamma^2 = q(x).$$

The identity (8) is equivalent to the identity

$$P(x) - (B_k x + C_k) Q_1(x) = q(x) P_1(x),$$

which owing to the arbitrariness of the polynomial $P_1(x)$ is, in its turn, equivalent to the condition that the polynomial on the left-hand

side is divisible by $q(x)$, i.e. by $x - \alpha$ and $x - \alpha^*$; but for this purpose it is necessary and sufficient that

$$P(\alpha) - (B_k \alpha + C_k) Q_1(\alpha) = P(\alpha^*) - (B_k \alpha^* + C_k) Q_1(\alpha^*) = 0,$$

or

$$B_k \alpha + C_k = \frac{P(\alpha)}{Q_1(\alpha)},$$

$$B_k \alpha^* + C_k = \frac{P(\alpha^*)}{Q_1(\alpha^*)}.$$

Hence we have a system of two equations of first degree with determinant $\alpha - \alpha^* = 2i\gamma \neq 0$ for the evaluation of the unknowns B_k and C_k , and we can therefore always determine these two numbers uniquely; it can readily be seen that in this case the expressions obtained for B_k and C_k depend symmetrically on α and α^* and they are therefore real. This proves lemma 2 completely.

If $k > 1$, then, as with a real root, the last fraction on the right-hand side of the identity (8) has the same form as the initial fraction on the left-hand side. We can therefore apply the same lemma. Continuing this process we find, as before, that if the polynomial $Q(x)$ has a complex root $\alpha = \beta + i\gamma$ ($\gamma \neq 0$) of multiplicity k and if the polynomial $Q_1(x)$ is defined by the identity (7), then the following identity holds :

$$\frac{P(x)}{Q(x)} = \frac{B_k x + C_k}{\{q(x)\}^k} + \frac{B_{k-1}x + C_{k-1}}{\{q(x)\}^{k-1}} + \dots + \frac{B_1x + C_1}{q(x)} + \frac{P^*(x)}{Q_1(x)}, \quad (9)$$

where $q(x) = (x - \beta)^2 + \gamma^2$, $B_1, B_2, \dots, B_k, C_1, C_2, \dots, C_k$ are real numbers and $P^*(x)$ is a polynomial with real coefficients.

We shall now make the following general remarks in connection with the identities (6) and (9); if the left-hand side of either of these identities is a regular rational fraction, then the last fraction on the right-hand side will also be regular; this can readily be proved if we assume that the variable x increases indefinitely; we can then see that all terms of the polynomial except $P^*(x)/Q_1(x)$ tend to zero; it follows from the identity that the last fraction must also tend to zero and it is possible only if the fraction is regular.

We are now able to convert every regular rational fraction into the "canonical" form which is convenient for integration. The denominator $Q(x)$ of this fraction, like any other polynomial with

real coefficients has, in general, several different real roots $\alpha_1, \alpha_2, \dots, \alpha_r$ and several pairs of conjugate imaginary roots $\beta_1 \pm i\gamma_1, \beta_2 \pm i\gamma_2, \dots, \beta_s \pm i\gamma_s$; every real root α_m occurs a definite number of times k_m ($1 \leq m \leq r$) and every pair of imaginary roots $\beta_n \pm i\gamma_n$ is of multiplicity l_n ($1 \leq n \leq s$). We know from algebra that

$$\begin{aligned} Q(x) &= a(x - \alpha_1)^{k_1} (x - \alpha_2)^{k_2} \dots (x - \alpha_r)^{k_r} (x - \beta_1 - i\gamma_1)^{l_1} \times \\ &\quad \times (x - \beta_1 + i\gamma_1)^{l_1} \dots (x - \beta_s - i\gamma_s)^{l_s} (x - \beta_s + i\gamma_s)^{l_s} = \\ &= a \prod_{m=1}^r (x - \alpha_m)^{k_m} \prod_{n=1}^s [x - \beta_n + i\gamma_n]^{l_n}, \quad (10) \end{aligned}$$

where $a \neq 0$ is a constant.

Applying the formula (6) to the given function $P(x)/Q(x)$ we obtain :

$$\frac{P(x)}{Q(x)} = \frac{A_1^{(1)} k_1}{(x - \alpha_1)^{k_1}} + \frac{A_1^{(1)} k_1 - 1}{(x - \alpha_1)^{k_1 - 1}} + \dots + \frac{A_1^{(1)}_1}{x - \alpha_1} + \frac{P_1(x)}{Q_1(x)}, \quad (11)$$

where $A_1^{(1)}, A_2^{(1)}, \dots, A_{k_1}^{(1)}$ are constant real numbers,

$$Q_1(x) = a \prod_{m=2}^r (x - \alpha_m)^{k_m} \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]^{l_n},$$

and the last fraction on the right-hand side is regular.

But the fraction $P_1(x)/Q_1(x)$ is of the same type as the given fraction $P(x)/Q(x)$ and we can again expand it in accordance with the formula (6) applied to the root α_2 , say; then we obtain :

$$\frac{P_1(x)}{Q_1(x)} = \frac{A_2^{(2)} k_2}{(x - \alpha_2)^{k_2}} + \dots + \frac{A_2^{(2)}_1}{x - \alpha_2} + \frac{P_2(x)}{Q_2(x)}, \quad (12)$$

where

$$Q_2(x) = a \prod_{m=3}^r (x - \alpha_m)^{k_m} \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]^{l_n},$$

and the last fraction is again regular. Substituting (12) in (11) we obtain :

$$\frac{P(x)}{P(x)} = \frac{A^{(1)}_{k_1}}{(x-\alpha_1)^{k_1}} + \dots + \frac{A^{(1)}_1}{x-\alpha_1} + \frac{A^{(2)}_{k_2}}{(x-\alpha_2)^{k_2}} + \dots + \frac{A^{(2)}_1}{x-\alpha_2} + \frac{P_2(x)}{Q_2(x)}.$$

After repeating this process r times (for all r real roots α_m) we evidently obtain the identity

$$\begin{aligned} \frac{P(x)}{Q(x)} = & \frac{A^{(1)}_{k_1}}{(x-\alpha_1)^{k_1}} + \frac{A^{(1)}_{k_1-1}}{(x-\alpha_1)^{k_1-1}} + \dots + \frac{A^{(1)}_1}{x-\alpha_1} + \\ & + \frac{A^{(2)}_{k_2}}{(x-\alpha_2)^{k_2}} + \frac{A^{(2)}_{k_2-1}}{(x-\alpha_2)^{k_2-2}} + \dots + \frac{A^{(2)}_1}{x-\alpha_2} + \\ & + \dots + \\ & + \frac{A^{(r)}_{k_r}}{(x-\alpha_r)^{k_r}} + \frac{A^{(r)}_{k_r-1}}{(x-\alpha_r)^{k_r-1}} + \dots + \frac{A^{(r)}_1}{x-\alpha_r} + \frac{P^*(x)}{Q^*(x)}, \end{aligned} \tag{13}$$

where the last fraction is regular and its denominator

$$Q^*(x) = a \prod_{n=1}^s [(x-\beta_n)^2 + \gamma_n^2]^{l_n}$$

has only the imaginary roots $\beta_n \pm i\gamma_n$ of the initial denominator $Q(x)$. Therefore the “separation” of real roots as expressed by formula (6) is no longer applicable to the fraction $P^*(x)/Q^*(x)$. We shall now naturally apply the process of “separation” of imaginary roots as defined by formula (9) to this fraction. Similarly we shall in this case obtain the following expansion for real roots by applying formula (9) s times

$$\begin{aligned} \frac{P^*(x)}{Q^*(x)} = & \frac{B^{(1)}_{l_1}x + C^{(1)}_{l_1}}{[(x-\beta_1)^2 + \gamma_1^2]^{l_1}} + \dots + \frac{B^{(1)}_1x + C^{(1)}_1}{(x-\beta_1)^2 + \gamma_1^2} + \\ & + \frac{B^{(2)}_{l_2}x + C^{(2)}_{l_2}}{[(x-\beta_2)^2 + \gamma_2^2]^{l_2}} + \dots + \frac{B^{(2)}_1x + C^{(2)}_1}{(x-\beta_2)^2 + \gamma_2^2} + \\ & + \dots + \\ & + \frac{B^{(s)}_{l_s}x + C^{(s)}_{l_s}}{[(x-\beta_s)^2 + \gamma_s^2]^{l_s}} + \dots + \frac{B^{(s)}_1x + C^{(s)}_1}{(x-\beta_s)^2 + \gamma_s^2} + \frac{P^{**}(x)}{Q^{**}(x)}, \end{aligned}$$

where the last fraction on the right-hand side is regular; but since all roots of the initial denominator $Q(x)$ have been used and $Q^{**}(x)$ has

no other roots, we must necessarily have $P^{**}(x) \equiv 3$. Substituting the above expansion of the fraction $P^*(x)/Q^*(x)$ in (13), we obtain the final expansion for the initial fraction which we can write in the following condensed form :

$$\frac{P(x)}{Q(x)} = \sum_{m=1}^r \sum_{u=1}^{k_m} \frac{A_u^{(m)}}{(x - \alpha_m)^u} + \sum_{n=1}^s \sum_{v=1}^{l_n} \frac{B_v^{(n)}x + C_v^{(n)}}{[(x - \beta_n)^2 + \gamma_n^2]^v}. \quad (14)$$

We were trying to obtain this expansion of the regular rational fraction $P(x)/Q(x)$; we have proved that it is always possible; at the same time it is also unique : at all stages of the successive determination of the numbers $A_u^{(m)}$, $B_v^{(n)}$ and $C_v^{(n)}$ we have found that their determination is unique. However, the above method for successive determination of coefficients of the expansion (14) is not usually the simplest method. It is generally easier and more symmetrical to use the so-called method of "undefined coefficients". We write the expansion (14) with *undefined* $A_u^{(m)}$, $B_v^{(n)}$ and $C_v^{(n)}$ and, in disposing of all fractions, multiply both sides of this relation by $Q(x)$. As a result we obtain the given polynomial $P(x)$ on the left-hand side and on the right-hand side another polynomial whose coefficients, after comparison with similar terms, evidently contains the unknown numbers $A_u^{(m)}$, $B_v^{(n)}$ and $C_v^{(n)}$ and, as can readily be seen, are *linearly* dependent on these numbers.

Since the resulting equation should be an identity, the coefficients of similar powers of x on the right and left-hand sides should be equal. Comparing them with each other in pairs, we obtain a system of equations of first degree for the unknowns $A_u^{(m)}$, $B_v^{(n)}$ and $C_v^{(n)}$, with whose help these numbers can be determined; we know in advance that this problem has a unique solution. It can be readily seen that the number of equations of the system is equal to the number of the unknowns. In fact, let us assume that the power of the polynomial $Q(x)$ is equal to N . On multiplying both sides of the identity (14) by $Q(x)$ we evidently obtain a polynomial of degree $N-1$ on the right-hand side; on the left-hand side we have the polynomial $P(x)$ whose power is not greater than $N-1$, since the fraction P/Q is regular. And since a polynomial of degree $N-1$ has N coefficients, a comparison of coefficients on the right and left-hand sides gives us a system of N equations. On the other hand, the number of the numbers $A_u^{(m)}$ ($1 \leq m \leq r$, $1 \leq u \leq k_m$)

is equal to $\sum_{m=1}^r k_m$; similarly the number of the numbers $B_v^{(n)}$ is equal

to $\sum_{n=1}^s l_n$ and the same applies to the number of the numbers $C_v^{(n)}$.

Hence the total number of the unknowns is equal to

$$\sum_{m=1}^r k_m + 2 \sum_{n=1}^s l_n;$$

but the expansion (10) of the polynomial $Q(x)$ into linear factors shows that this number is exactly equal to the power N of the polynomial $Q(x)$. Hence the number of the unknowns is, in fact, always equal to the number of linear equations obtained.

Thus in order to write the expansion (14) irrespective of the methods by which it was obtained, it is always necessary to know all roots of the polynomial $Q(x)$ and their multiplicity. This is an algebraic problem which we cannot always solve, but we must nevertheless assume that this problem is solvable before proceeding with integration of the given rational fraction.

Example. The fraction

$$\frac{2x + 2}{(x - 1)(x^2 + 1)^2}$$

can, according to formula (14), be represented in the following form:

$$\frac{2x + 2}{(x - 1)(x^2 + 1)^2} = \frac{A}{x - 1} + \frac{B_1x + C_1}{(x^2 + 1)^2} + \frac{B_2x + C_2}{x^2 + 1};$$

multiplying both sides by $(x - 1)(x^2 + 1)^2$ and comparing similar terms on the right and left-hand sides we obtain:

$$2x + 2 = (A + B_2)x^4 + (C_2 - B_2)x^3 + (2A + B_1 + B_2 - C_2)x^2 + \\ + (C_1 - B_1 + C_2 - B_2)x + (A - C_1 - C_2).$$

Comparing the corresponding coefficients with one another on the right and left-hand sides we obtain the following system of equations:

$$\begin{aligned} A + B_2 &= 0, & C_2 - B_2 &= 0, \\ 2A + B_1 + B_2 - C_2 &= 0, \\ C_1 + C_2 - B_1 - B_2 &= 2, \\ A - C_1 - C_2 &= 2; \end{aligned}$$

this system can be readily solved and we obtain :

$$A = 1, \quad B_1 = -2, \quad C_1 = 0, \quad B_2 = -1, \quad C_2 = -1.$$

Therefore

$$\frac{2x+2}{(x-1)(x^2+1)^2} = \frac{1}{x-1} - \frac{2x}{(x^2+1)^2} - \frac{x+1}{x^2+1}.$$

§ 60. Integration of simple fractions

Formula (14) introduced in § 59 shows that integration of any regular (and hence also any irregular) rational fraction involves integration of a series of rational fractions of a special kind which we shall call *simple fractions*. Simple fractions can be divided into two types : fractions of the type

$$\frac{A}{(x-\alpha)^u},$$

where A and α are constant real numbers and u a constant natural number, are called fractions of the *first kind*, and fractions of the type

$$\frac{Bx+C}{[(x-\beta)^2+\gamma^2]^v},$$

where B , C , β and α are constant real numbers and v is a constant natural number, are called fractions of the *second kind*. In this paragraph we shall learn to find primitives of functions for all simple fractions of either kind. With the help of formula (14) we shall then be able to say that we have learnt to integrate every rational function.

1°. **Simple fractions of the first kind.** We directly find that

$$\int \frac{A dx}{x-\alpha} = A \ln |x-\alpha| + H,$$

where H is the constant of integration. Similarly when $u > 1$, we directly obtain

$$\begin{aligned} \int \frac{A dx}{(x-\alpha)^u} &= \int A (x-\alpha)^{-u} dx = \frac{A}{-u+1} (x-\alpha)^{-u+1} + H = \\ &= \frac{-A}{(u-1)(x-\alpha)^{u-1}} + H. \end{aligned}$$

This evidently concludes integration of fractions of the first kind. We can see that the primitives obtained as a result of this process are

either other simple fractions of the first kind (for $u > 1$) or logarithmic functions for $u = 1$.

2°. Simple fractions of the second kind. Let us at first assume that $v = 1$, i.e. we are dealing with a simple fraction of the following type

$$\frac{Bx + C}{(x - \beta)^2 + \gamma^2}.$$

The substitution $x = \beta + \gamma y$ ($y = (x - \beta) / \gamma$, $dx = \gamma dy$) gives :

$$\begin{aligned} \int \frac{Bx + C}{(x - \beta)^2 + \gamma^2} dx &= \int \frac{B(\beta + \gamma y) + C}{\gamma^2(1 + y^2)} \gamma dy = \\ &= \frac{B}{2} \int \frac{2y dy}{1 + y^2} + \frac{B\beta + C}{\gamma} \int \frac{dy}{1 + y^2} = \\ &= \frac{B}{2} \ln(1 + y^2) + \frac{B\beta + C}{\gamma} \arctan y + H = \\ &= \frac{B}{2} \ln \left\{ 1 + \left(\frac{x - \beta}{\gamma} \right)^2 \right\} + \frac{B\beta + C}{\gamma} \arctan \left(\frac{x - \beta}{\gamma} \right) + H. (1) \end{aligned}$$

Let us now assume that v is an arbitrary natural number. In this case the same substitution $x = \beta + \gamma y$ gives :

$$\begin{aligned} \int \frac{Bx + C}{[(x - \beta)^2 + \gamma^2]^v} dx &= \int \frac{B(\beta + \gamma y) + C}{\gamma^{2v}(1 + y^2)^v} \gamma dy = \\ &= \frac{B}{2\gamma^{2v-2}} \int \frac{2y dy}{(1 + y^2)^v} + \frac{B\beta + C}{\gamma^{2v-1}} \int \frac{dy}{(1 + y^2)^v}. \end{aligned}$$

Here the first primitive on the right-hand side can again be found directly :

$$\int \frac{2y dy}{(1 + y^2)^v} = - \frac{1}{(v - 1)(1 + y^2)^{v-1}} + H$$

(we assume that $v > 1$, since we have already considered above the case when $v = 1$). Hence in order to solve the problem completely it only remains to find the primitive

$$I_v = \int \frac{dy}{(1 + y^2)^v},$$

where v is an arbitrary natural number (for the moment we only know that $I_1 = \arctan y = H$). With this view we shall now deduce the reduction formula which expresses I_{v+1} in terms of I_v for every natu-

ral v . Hence by knowing I_1 we can find in succession I_2, I_3 , etc. and generally I_v for every v .

We have:

$$I_{v+1} = \int \frac{dy}{(1+y^2)^{v+1}} = \int \frac{(1+y^2) - y^2}{(1+y^2)^{v+1}} dy = \\ = I_v - \frac{1}{2} \int y \frac{2y dy}{(1+y^2)^{v+1}}. \quad (2)$$

The last primitive on the right-hand side can be integrated by parts; knowing that

$$\int \frac{2y dy}{(1+y^2)^{v+1}} = -\frac{1}{v(1+y^2)^v} + H,$$

we obtain:

$$\int y \frac{2y dy}{(1+y^2)^{v+1}} = -\frac{y}{v(1+y^2)^v} + \frac{1}{v} \int \frac{dy}{(1+y^2)^v} = \\ = -\frac{y}{v(1+y^2)^v} + \frac{1}{v} I_v,$$

and the equation (2) therefore gives us

$$I_{v+1} + \frac{2v-1}{2v} I_v + \frac{y}{2v(1+y^2)^v}, \quad (3)$$

and in particular

$$I_2 = \frac{1}{2} I_1 = \frac{y}{2(1+y^2)} = \frac{1}{2} \arctan y + \frac{y}{2(1+y^2)} + H,$$

etc. Formula (3) is the required reduction formula; thus by deducing this formula we can integrate simple fractions of the second kind.

A comparison of results shows that the functions obtained by integration of simple fractions (and therefore also by integrating all rational fractions) can either be logarithms and arctangents or rational functions. This also confirms the above expressed view that primitives of all rational functions are always elementary functions.

Let us make one more interesting remark. At the beginning of our study of integration we have already drawn attention to the fact

that the functions $\ln x$ and $\arctan x$ are primitives of very simple rational fractions :

$$\int \frac{dx}{x} = \ln x + C, \quad \int \frac{dx}{1+x^2} = \arctan x + C.$$

Now that we have completely developed the theory of integrating rational functions, we can see that no matter how complicated the given rational functions be, their primitives can always be expressed in terms of the following two transcendental functions: $\ln x$ and $\arctan x$.

Numerous exercises for integrating rational functions by means of the method of indefinite coefficients can be found in the Problem Book by B.P. Demidovich, Section III, Nos. 174-190; the teacher should select about 3 or 4 problems.

§ 61. Ostrogradski's Method

We have seen in earlier paragraphs that integration of rational fractions, where the roots of the denominator are known, does not cause undue difficulties, although it is often connected with rather lengthy calculations. M.V. Ostrogradski found an ingenious general method which can often simplify and shorten these calculations. To explain this method we shall have to revert to arguments used in the last two paragraphs.

Let us assume again that $P(x)/Q(x)$ is a regular rational fraction and

$$Q(x) = a \prod_{m=1}^r (x - \alpha_m)^{k_m} \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]^{l_n}. \quad (1)$$

We have seen earlier that it is possible to expand the fraction $P(x)/Q(x)$ uniquely into simple fractions of the first and second kind by using the expansion (14) § 59 used for integration of the given fraction. In this case the position is as follows :

The fraction of the first kind

$$\frac{A}{(x - \alpha)^u}$$

gives on integration natural logarithms for $u = 1$ and rational functions of the type

$$\int \frac{A dx}{(x - \alpha)^u} = - \frac{A}{(u - 1)(x - \alpha)^{u-1}} + H \quad (2)$$

for $u > 1$.

The position is slightly more complicated with fractions of the second kind

$$\frac{Bx + C}{[(x - \beta)^2 + \gamma^2]^v}.$$

Assuming that $x = \beta + \gamma y$ we obtain for $v > 1$:

$$\int \frac{Bx + C}{[(x - \beta)^2 + \gamma^2]^v} dx = \frac{\lambda_v}{(1 + y^2)^{v-1}} + \mu_v I_v, \quad (3)$$

where

$$I_v = \int \frac{dy}{(1 + y^2)^v},$$

and λ_v and μ_v are constants. On the other hand successive application of reduction formula (3) § 60 evidently enables us to represent the primitive I_v as a sum

$$I_v = \nu_v I_1 + \frac{L(y)}{(1 + y^2)^{v-1}},$$

where ν_v is a constant, $L(y)$ is a polynomial and the last fraction is regular. Substituting this expression in formula (3) and returning on the right-hand side from the variable y to the variable x we readily obtain:

$$\int \frac{Bx + C}{[(x - \beta)^2 + \gamma^2]^v} dx = \frac{R(x)}{[(x - \beta)^2 + \gamma^2]^{v-1}} + \sigma_v \int \frac{dx}{(x - \beta)^2 + \gamma^2}, \quad (4)$$

where $R(x)$ is a polynomial, σ_v is a constant and the first fraction on the right-hand side is regular. This is the position for $v > 1$; when $v = 1$ we have formula (1) § 60 in which there are no rational terms on the right-hand side.

We now have a clear picture of the primitive of the fraction P/Q when expanded in accordance with the expansion (14) § 59. We can see ((2) and (4)) that the terms of this expansion in which $u > 1$ or $v > 1$ give on integration regular rational fractions with corresponding denominators

$$(x - \alpha_m)^{u-1}, [(x - \beta_n)^2 + \gamma_n^2]^{v-1}.$$

On adding all these regular fractions we obtain another regular fraction

$$\frac{P_1(x)}{Q_1(x)},$$

whose denominator is evidently equal to

$$Q_1(x) = \prod_{m=1}^r (x - \alpha_m)^{k_m-1} \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]^{l_n-1}. \quad (5)$$

This is the rational part of the integral of the given fraction P/Q . The second transcendental part will evidently consist of: (a) primitives of those terms of the expansion (14) § 59 in which $u = 1$ and $v = 1$, and (b) primitives of the second kind of the second term in formula (4). In all these cases the integrand belongs to one of the following types :

$$\frac{A}{x - \alpha}, \quad \frac{Bx + C}{- \beta^2 + \gamma^2};$$

the sum of these integrands will therefore be a regular rational fraction

$$\frac{P_2(x)}{Q_2(x)},$$

where

$$Q_2(x) = \prod_{m=1}^r (x - \alpha_m) \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]. \quad (6)$$

We thus obtain Ostrogranski's remarkable formula

$$\int \frac{P(x)}{Q(x)} dx = \frac{P_1(x)}{Q_1(x)} + \int \frac{P_2(x)}{Q_2(x)} dx, \quad (7)$$

where the first and second terms on the right-hand side represent the rational and transcendental parts of the primitive respectively. $Q_1(x)$ and $Q_2(x)$ are respectively determined from the formulae (5) and (6) and the fractions $P_1(x) / Q_1(x)$ are regular.

The most remarkable feature of this expansion is due to the fact that it can be obtained by rational deduction *without knowing the roots of the polynomial* $Q(x)$. In fact, we know from algebra that a root

of multiplicity $k \geq 1$ of the polynomial $Q(x)$ is a root of multiplicity $k - 1$ of the polynomial $Q'(x)$; if we therefore assume that

$$Q(x) = a \prod_{m=1}^r (x - \alpha_m)^{k_m} \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]^{l_n},$$

then

$$Q'(x) = \prod_{m=1}^r (x - \alpha_m)^{k_m-1} \prod_{n=1}^s [(x - \beta_n)^2 + \gamma_n^2]^{l_n-1} R(x) = Q_1(x) R(x),$$

where the polynomials $Q(x)$ and $R(x)$ have no common root. This shows that the polynomial $Q_1(x)$ is the *greatest common divisor* of the polynomials $Q(x)$ and $Q'(x)$ and can therefore be obtained by the usual method for finding the greatest common divisor of two polynomials, *i.e.* by successive division. And since the formulae (1), (5) and (6) give :

$$Q(x) = a Q_1(x) Q_2(x),$$

therefore knowing $Q(x)$ and $Q_1(x)$ we can find the polynomial $Q_2(x)$ by elementary operations. Finally to obtain the polynomials $P_1(x)$ and $P_2(x)$ we can differentiate the equation (7) :

$$\frac{P(x)}{Q(x)} = \frac{Q_1(x) P_1'(x) - P_1(x) Q_1'(x)}{Q_1^2(x)} + \frac{P_2(x)}{Q_2(x)}. \quad (8)$$

According to formula (5) each root λ of the polynomial $Q_1(x)$ is a root of the polynomial $Q(x)$ and it follows from formula (6) that it is also a root of the polynomial $Q_2(x)$. If $Q_1(x)$ contains the binomial $x - \lambda$ of power $k > 0$, then $Q_1'(x)$ contains it also, but its power is $k - 1$, and it appears in $Q_2(x)$ in the first degree; therefore the product $Q_1'(x) Q_2(x)$ contains $x - \lambda$ of the same power k as the polynomial $Q_1(x)$; and since the same also applies to any root λ of the polynomial $Q_1(x)$, $Q_1'(x) Q_2(x)$ is divisible by $Q_1(x)$ without remainder, *i.e.*

$$Q_1'(x) Q_2(x) = Q_1(x) S(x),$$

where $S(x)$ is a polynomial. We therefore obtain :

$$\begin{aligned} \frac{Q_1(x) P_1'(x) - P_1(x) Q_1'(x)}{Q_1^2(x)} &= \frac{Q_2(x) Q_1(x) P_1'(x) - Q_2(x) P_1(x) Q_1'(x)}{Q_2(x) Q_1^2(x)} \\ &= \frac{Q_1(x) [Q_2(x) P_1'(x) - P_1(x) S(x)]}{Q_2(x) Q_1^2(x)} = \frac{Q_2(x) P_1'(x) - P_1(x) S(x)}{Q_1(x) Q_2(x)}, \end{aligned}$$

and after multiplying by $Q(x) = a Q_1(x) Q_2(x)$ the expansion (8) gives :

$$P(x) = a [Q_2(x) P'_1(x) - P_1(x) S(x)] + a P_2(x) Q_1(x). \quad (9)$$

In this expansion the polynomials $P(x)$, $Q_1(x)$, $Q_2(x)$ and $S(x)$ are known to us; the highest possible powers of the polynomials $P_1(x)$ and $P_2(x)$ which we are trying to find is determined by the regularity of the fractions $P_1(x)/Q_1(x)$ and $P_2(x)/Q_2(x)$. Hence the polynomials P_1 and P_2 can be readily obtained from the relation (9) by the method of indefinite coefficients. It can be readily seen that in this case the number of the unknowns coincides with the number of equations obtained and the solution of this system is guaranteed by the expansion (8).

Hence all elements in Ostrogradski's formula can, in fact, be deduced rationally and their deduction requires no knowledge of the roots of the denominator of the given fraction. Thus, without knowing these roots, we can find the rational part of the primitive of the given rational fraction.

For exercises on Ostrogradski's method cf. Problem Book B. P. Demidovich, Section III, Nos. 191-193.

CHAPTER XVII

INTEGRATION OF SIMPLE RATIONAL AND TRANSCENDENTAL FUNCTIONS

We have seen in the last chapter that all rational functions have elementary primitives, and we have found a general method for evaluating these primitives. However, as soon as we go beyond the class of rational functions, existence of elementary primitives is no longer a rule than an exception; therefore we can no longer construct general theories in the way we did in Chapter XVI. Nevertheless, algebraic irrational functions and transcendental functions contain fairly wide classes which give elementary functions on integration; these functions include rather simple functions which occur very often in applications; the methods available for finding primitives of these functions are rather instructive, and we shall consider several important functions of this kind in this chapter. In integrating irrational and transcendental functions the so-called *rationalisation method* is of great importance if it is required to transform the variable of integration so as to convert the integrand into a rational function; if this can be done, we can regard our problem solved in principle, since we can always integrate rational functions.

§ 62. Integration of functions of the type $R \left(x, \sqrt[n]{\frac{ax+b}{cx+d}} \right)$.

When we go beyond the region of rational functions, we find among simple functions some functions which besides rational operations involve extraction of a root. As an example of a very general function of this type we can take an arbitrary rational function $f(x) = P(x)/Q(x)$ from which a root of degree n is to be extracted (where $P(x)$ and $Q(x)$ are polynomials); we must then take an arbitrary rational function $R(x, y)$ of the variables x and

$$y = \sqrt[n]{\frac{P(x)}{Q(x)}};$$

hence the following functions can be regarded as simple primitives of irrational algebraic functions

$$\int R \left\{ x, \sqrt[n]{\frac{P(x)}{Q(x)}} \right\} dx,$$

where $P(x)$ and $Q(x)$ are polynomials, $n > 1$ is an arbitrary natural number and $R(x, y)$ is an arbitrary rational function of two variables.

However, among primitives of the type (1) there are very few which satisfy even the simplest conditions made with regard to the number n and the polynomials P and Q which can be expressed in terms of elementary functions. In this paragraph we shall consider the case when P and Q are linear binomials (and the number n is arbitrary); we shall see that primitives of this type are elementary functions which can be readily found.

Thus we are trying to find the primitive

$$\int R \left\{ x, \sqrt[n]{\frac{ax+b}{cx+d}} \right\} dx,$$

where a, b, c and d are constants, and n is an arbitrary natural number. If we assume that

$$\sqrt[n]{\frac{ax+b}{cx+d}} = t, \quad (2)$$

then

$$\frac{ax+b}{cx+d} = t^n, \quad x = \frac{dt^n - b}{n - ct^n} = \varphi(t), \quad dx = \varphi'(t) dt,$$

and therefore

$$\int R \left\{ x, \sqrt[n]{\frac{ax+b}{cx+d}} \right\} dx = \int R \{ \varphi(t), t \} \varphi'(t) dt.$$

Since the function $\varphi(t)$ (and therefore also its derivative $\varphi'(t)$) is rational, therefore, on the right-hand side we are dealing with a primitive of a rational function which can be expressed in terms of an elementary function of t ; replacing t by its expression (2) in terms of x we find the expression for the required primitive in terms of an elementary function of x .

Example. We must find the primitive

$$\int \frac{dx}{\sqrt[3]{1+x} - \sqrt[4]{1+x}}.$$

Since both radicals in the denominator are integral positive powers of the same radical $^{12}\sqrt{1+x}$, this primitive belongs to the class which we have just considered above ($n = 12$, $a = b = d = 1$; $C = 0$).

Assuming that

$$^{12}\sqrt{1+x} = t,$$

we obtain $x = t^{12} - 1$, $dx = 12 t^{11} dt$, $^{12}\sqrt{1+x} = t$, $^{12}\sqrt{1+x} = t^3$, and the given primitive is transformed into

$$12 \int \frac{t^{11} dt}{t^4 - t^3} = 12 \int \frac{t^8 dt}{t-1}.$$

Rationalisation is thus complete ; we obtain

$$\begin{aligned} 12 \int \frac{t^8 dt}{t-1} &= 12 \int \frac{t^8 - 1}{t-1} dt + 12 \int \frac{dt}{t-1} = \\ &= 12 \int (t^7 + t^6 + t^5 + t^4 + t^3 + t^2 + t + 1) dt + 12 \int \frac{dt}{t-1} = \\ &= 12 \left\{ \frac{t^8}{8} + \frac{t^7}{7} + \frac{t^6}{6} + \frac{t^5}{5} + \frac{t^4}{4} + \frac{t^3}{3} + \frac{t^2}{2} + t \right\} + 12 \ln |t-1| + C. \end{aligned}$$

Substituting here $t = ^{12}\sqrt{1+x}$ we obtain the required expression for the given primitive in terms of the initial variable x .

For further examples to § 62 *cf.* Problem Book by B. P. Demidovich, Section III, Nos. 211, 212, 215, 217.

§ 63. Integration of functions of the type $R(x, \sqrt{ax^2 + bx + c})$

If at least one of the polynomials P and Q in § 62 were of a higher degree than the first, then integration in terms of elementary functions would have been possible only in a few isolated cases. We shall now consider a case which often occurs in applications when $n = 2$, $Q(x) = 1$ and $P(x)$ is a trinomial of second degree $ax^2 + bx + c$; we are therefore dealing here with a primitive of the type

$$\int R(x, \sqrt{ax^2 + bx + c}) dx,$$

where $R(x, y)$ is, as before, an arbitrary rational function of two variables. We will now show that it is always possible to rationalise such a primitive and this primitive must therefore be an elementary function. The transformation of the variable of integration necessary for this rationalisation differs in each case.

1°. If the roots α and β of the trinomial $ax^2 + bx + c$ are real, we have (assuming that $x > \alpha$)

$$\sqrt{ax^2 + bx + c} = \sqrt{a(x-\alpha)(x-\beta)} = (x-\alpha) \sqrt{\frac{a(x-\beta)}{x-\alpha}};$$

hence the integrand depends rationally on x and on the radical

$$\sqrt{\frac{a(x-\beta)}{x-\alpha}},$$

and this brings us to the case considered in § 62; we know that rationalisation can be achieved by replacing the variable

$$\sqrt{\frac{a(x-\beta)}{x-\alpha}} = t.$$

2°. If the roots of the trinomial $ax^2 + bx + c$ are imaginary, this trinomial preserves the same sign for all values of x ; we are naturally assuming that it is always positive; otherwise the value of the radical would be imaginary for every value of x and the problem would become void. In particular, by assuming that $x = 0$ we can see that in this case we must necessarily have $c > 0$ (the method which we are now going to describe always leads to the desired result for $c > 0$ irrespective of whether the roots of the trinomial are real or imaginary). Assuming that

$$\frac{\sqrt{ax^2 + bx + c} - \sqrt{c}}{x} = t,$$

we obtain

$$ax^2 + bx + c = (tx + \sqrt{c})^2 = t^2x^2 + 2\sqrt{c}tx + c,$$

$$ax + b = t^2x + 2\sqrt{c}t,$$

$$x = \frac{b - 2\sqrt{c}t}{t^2 - a} = \varphi(t),$$

$$dx = \varphi'(t) dt,$$

$$\sqrt{ax^2 + bx + c} = tx + \sqrt{c} = t\varphi(t) + \sqrt{c},$$

and therefore

$$\int R(x, \sqrt{ax^2 + bx + c}) dx = \int R\{\varphi(t), t\varphi(t) + \sqrt{c}\} \varphi'(t) dt.$$

Owing to the fact that the function $\varphi(t)$ (and therefore also its derivative $\varphi'(t)$) is rational, we succeed in rationalising the given primitive.

In both cases the methods for transformation of the variable of integration were indicated by L. Euler and they are therefore known as *Euler's substitutions*.

Example 1. In the primitive

$$I = \int \frac{dx}{\sqrt{x^2 - a^2}}$$

($a > 0$, $|x| > a$), the roots of the polynomial $x^2 - a^2 = (x - a)(x + a)$ are real. Applying Euler's first substitution

$$\sqrt{\frac{x-a}{x+a}} = t$$

and assuming that $x > a$ we obtain :

$$\frac{x-a}{x+a} = t^2, \quad x = a \frac{1+t^2}{1-t^2}, \quad dx = \frac{4at}{(1-t^2)^2} dt,$$

$$x+a = \frac{2a}{1-t^2}, \quad \frac{1}{\sqrt{x^2-a^2}} = \frac{1}{\sqrt{\frac{x-a}{x+a}(x+a)}} = \frac{1-t^2}{2at},$$

and consequently

$$I = 2 \int \frac{dt}{1-t^2} = \int \frac{dt}{1+t} + \int \frac{dt}{1-t} = \ln \left| \frac{1+t}{1-t} \right| + C.$$

But

$$\begin{aligned} \frac{1+t}{1-t} &= \frac{\sqrt{x+a} - \sqrt{x-a}}{\sqrt{x+a} - \sqrt{x-a}} = \frac{1}{2a} (\sqrt{x+a} + \sqrt{x-a})^2 \\ &= \frac{1}{a} (x + \sqrt{x^2 - a^2}), \end{aligned}$$

and therefore

$$I = \ln (x + \sqrt{x^2 - a^2}) + C;$$

when $x < -a$ we obtain similarly :

$$I = \ln |x + \sqrt{x^2 - a^2}| + C.$$

Example 2. In the primitive

$$I = \int \frac{dx}{\sqrt{x^2 + a^2}},$$

$a^2 < 0$ and we can therefore apply Euler's second substitution :

$$\frac{\sqrt{x^2 + a^2} - a}{x} = t, \quad \sqrt{x^2 + a^2} = xt + a,$$

and we obtain :

$$\begin{aligned} x &= \frac{2at}{1-t^2}, & dx &= \frac{2a(1+t^2)}{(1-t^2)^2} dt, \\ \sqrt{x^2 + a^2} &= \frac{2at^2}{1-t^2} + a = a \frac{1+t^2}{1-t^2}, \\ \sqrt{x^2 + a^2} + x &= a \frac{1+t}{1-t}, \end{aligned}$$

and therefore

$$I = 2 \int \frac{dt}{1-t^2} = \ln \left| \frac{1+t}{1-t} \right| + C = \ln (x + \sqrt{x^2 + a^2}) + C.$$

Example 3. To the primitive

$$I = \int \frac{dx}{\sqrt{a^2 - b^2}}$$

either of the two Euler's substitutions can be applied, However, in this case it would be simplest to substitute $x = at$; we obtain

$$I = \int \frac{dx}{\sqrt{1-t^2}} = \arcsin t + C = \arcsin \frac{x}{a} + C.$$

For further examples to § 63 cf. Problem Book by B. P. Demidovich, Section III, Nos. 219-222, 245-247.

§ 64. Primitives of binomial differentials

We shall now consider integration of a special class of algebraic functions; integrals of this type frequently occur in applications; however, the historical significance of this method is mainly due to the fact that it is one of the rare cases of the problem where integrals of the given class are elementary but not the functions of that class.

A *binomial differential* is an expression of the type

$$x^\alpha (a + bx^\beta)^\gamma dx,$$

where all three indices α , β and γ are rational and a and b are

arbitrary real numbers. We shall find under what conditions the primitive

$$I = \int x^\alpha (a + bx^\beta)^\gamma dx$$

can be an elementary function.

Let us assume that $x^\beta = t$ so that ^{*)}

$$x = t^{\frac{1}{\beta}}, \quad dx = \frac{1}{\beta} t^{\frac{1}{\beta} - 1} dt.$$

We thus obtain

$$I = \frac{1}{\beta} \int t^{\frac{\alpha+1}{\beta} - 1} (a + bt)^\gamma dt. \quad (1)$$

We shall now see that *if at least one of the three numbers $(\alpha+1)/\beta$, γ and $(\alpha+1)/\beta + \gamma$ is an integer, then I represents an elementary function.* We shall also give a method for finding this function.

1° Let γ be an integer. In that case the integrand in the primitive (1) depends rationally on $t^{\frac{\alpha+1}{\beta}}$ and t ; if

$$\frac{\alpha+1}{\beta} = \frac{m}{n},$$

where m and n are integers ($n > 0$), this integrand has the form $R(t, \sqrt[n]{t})$, where $R(x, y)$ is a rational function of two variables. The primitive I has the form considered in § 62 and it can therefore be expressed in terms of elementary functions.

2°. Let the number $(\alpha+1)/\beta$ be an integer. In that case the integrand in (1) depends rationally on t and $(a + bt)^\gamma$; if $\gamma = p/q$, where p and $q < 0$ are integers, then this integrand has the form $R(t, \sqrt[q]{a+bt})$ and we have again a primitive of the type considered in § 62).

3°. Finally, let us assume that $(\alpha+1)/\beta + \gamma$ is an integer. The integrand in (1) can then be written in the form

$$\left(\frac{a+bt}{t}\right)^\gamma t^{\frac{\alpha+1}{\beta} - 1 + \gamma}$$

^{*)} We can assume that $\beta \neq 0$, for the case $\beta = 0$ is evidently trivial.

and it therefore depends rationally on t and $(a + bt) / t$ if $\gamma = p/q$; If p and $q > 0$ are integers, then the integrand has the form

$$R\left(t, \sqrt[q]{\frac{a + bt}{t}}\right),$$

and we have again the conditions considered in § 62.

Thus our proposition is fully proved. P. L. Chebyshev has shown that the above conditions include all cases when the primitive of a binomial differential is an elementary function; if a and b are non-zero and none of the three numbers $(\alpha+1)/\beta$, γ and $(\alpha+1)/\beta + \gamma$ is an integer, then the primitive is never an elementary function. Unfortunately the proof of Chebyshev's remarkable theorem is too complicated to be given here.

For problems to § 64 cf. Problem Book by B. P. Demidovich, Section III, Nos. 252, 253, 260.

§ 65. Integration of trigonometrical differentials

We shall now integrate some classes of transcendental functions and at first consider functions which depend rationally on the trigonometrical functions $\sin x$, $\cos x$, $\tan x$, $\cot x$, $\sec x$ and $\operatorname{cosec} x$. Owing to the fact that all these functions can be expressed rationally in terms of $\sin x$ and $\cos x$, we are obviously dealing with functions of the type $R(\sin x, \cos x)$, where $R(x, y)$ is a rational function of two variables.

The primitive

$$I = \int R(\sin x, \cos x) dx \quad (1)$$

always represents an elementary function. To prove this, it is sufficient to introduce the following quantity as the new variable of integration

$$\tan\left(\frac{x}{2}\right) = t.$$

In this case by only taking values of x in the interval $-\pi/2 < x < \pi/2$ we obtain

$$\begin{aligned} x &= 2 \arctan t, & dx &= \frac{2dt}{1+t^2}, \\ \sin x &= \frac{2t}{1+t^2}, & \cos x &= \frac{1-t^2}{1+t^2} \end{aligned}$$

and

$$I = \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \frac{2dt}{1+t^2}$$

is the primitive of the rational function.

Example 1. Find the primitive

$$I = \int \frac{dx}{1 - \lambda^2 \cos^2 x},$$

where λ^2 is an arbitrary positive number. We shall consider separately the case when $\lambda^2 < 1$, $\lambda^2 > 1$ and $\lambda^2 = 1$.

1) If $\lambda^2 < 1$, we can assume that $1 - \lambda^2 = \alpha^2$, $1 + \lambda^2 = \beta$. Substituting in (2) we obtain :

$$\begin{aligned} I &= \int \frac{2dt}{1+t^2} \cdot \frac{1+t^2}{1+t^2 - \lambda^2(1-t^2)} = \int \frac{2dt}{\alpha^2 + \beta^2 t^2} = \frac{2}{\alpha\beta} \arctan \frac{\beta t}{\alpha} + \\ &+ C = \frac{2}{\sqrt{1-\lambda^4}} \arctan \left(\sqrt{\frac{1+\lambda^2}{1-\lambda^2}} \tan \frac{x}{2} \right) + C. \end{aligned}$$

2) If $\lambda^2 > 1$, we assume that $1 - \lambda^2 = -\alpha^2$, $1 + \lambda^2 = \beta$. Substituting in (2) we obtain :

$$\begin{aligned} I &= \int \frac{2dt}{\beta^2 t^2 - \alpha^2} = \frac{1}{\alpha\beta} \ln \left| \frac{\beta t - \alpha}{\beta t + \alpha} \right| + C = \\ &= \frac{1}{\alpha\beta} \ln \left| \frac{\beta \tan \frac{x}{2} - \alpha}{\beta \tan \frac{x}{2} + \alpha} \right| + C = \frac{1}{\sqrt{\lambda^4 - 1}} \ln \left| \frac{\sqrt{\lambda^2 + 1} \tan \frac{x}{2} - \sqrt{\lambda^2 - 1}}{\sqrt{\lambda^2 + 1} \tan \frac{x}{2} + \sqrt{\lambda^2 - 1}} \right| + C. \end{aligned}$$

3) If $\lambda^2 = 1$, the same substitution gives :

$$I = \int \frac{dt}{t^2} = -\frac{1}{t} + C = -\cot \frac{x}{2} + C.$$

As a result of the substitution (2) the primitive (1) is rationalised in every case, which is very important as it shows that all primitives of the kind (1) represent elementary functions. However, this substitution is often rather difficult in practice and can be replaced by other simpler transformations of the independent variable. Many cases can be quoted when primitives of the type (1) can be rationalised by simpler substitutions like $t = \sin x$, $t = \cos x$ or $t = \tan x$. We shall consider some such cases.

1°. If the function $R(x, y)$ is odd with respect to y (i.e. it only changes its sign as y is replaced by $-y$), the primitive (1) can be rationalised by means of the transformation $\sin x = t$. In fact, in this case the function

$$\frac{R(\sin x, \cos x)}{\cos x} \quad (3)$$

does not change when $\cos x$ is replaced by $-\cos x$ and therefore *) contains only the square of $\cos x$; but $\cos^2 x = 1 - \sin^2 x$, so that the function (3) depends rationally on $\sin x = t$. We therefore have

$$\int R(\sin x, \cos x) dx = \int \frac{R(\sin x, \cos x)}{\cos x} \cos x dx = \int R^*(t) dt,$$

where $R^*(t)$ is some rational function of t .

2°. It can be similarly shown that if the function $R(x, y)$ is odd with respect to x , the primitive (1) can be rationalised by means of the transformation $\cos x = t$.

3°. Finally if $R(x, y) \equiv R(-x, -y)$, the primitive (1) can be rationalised by means of the transformation $\tan x = t$. In fact, replacing everywhere $\sin x$ by $\cos x \tan x$ we obtain a rational function $R_1(\tan x, \cos x)$ of $\tan x$ and $\cos x$ so that

$$R(\sin x, \cos x) \equiv R_1(\tan x, \cos x),$$

and therefore

$$R(-\sin x, -\cos x) \equiv R_1(\tan x, -\cos x).$$

Since the left-hand sides of these two identities coincide, therefore

$$R_1(\tan x, \cos x) \equiv R_1(\tan x, -\cos x),$$

*) We are using here the algebraic theorem: if $R(z)$ is a rational function of z and $R(-z) = R(z)$, then $R(z)$ is a rational function of z^2 . Proof. We are given that $R(z) = R(-z) = \frac{1}{2} [R(z) + R(-z)]$; if $R(z)$ is a polynomial, the right-hand side is evidently a polynomial with respect to z^2 ; in general, let us assume that $R(z)$ is equal to $P(z) / Q(z)$ so that

$$R(z) + R(-z) = \frac{P(z) Q(-z) + Q(z) P(-z)}{Q(z) Q(-z)};$$

both the numerator and the denominator are polynomials and they evidently do not change when z is replaced by $-z$; therefore, in accordance with our proof they are polynomials with respect to z^2 .

i.e. the function R_1 does not change when $\cos x$ is replaced by $-\cos x$ and therefore it only contains the square of $\cos x$:

$$R_1(\tan x, \cos x) = R_2(\tan x, \cos^2 x)$$

and hence also

$$R(\sin x, \cos x) = R_2(\tan x, \cos^2 x).$$

Assuming that $\tan x = t$ we obtain :

$$\cos^2 x = \frac{1}{1+t^2}, \quad x = \arctan t, \quad dx = \frac{dt}{1+t^2},$$

and consequently

$$\begin{aligned} I &= \int R(\sin x, \cos x) dx = \int R_2(\tan x, \cos^2 x) dx = \\ &= \int R_2\left(t, \frac{1}{1+t^2}\right) \frac{dt}{1+t^2}, \end{aligned}$$

and the primitive is, in fact, rationalised.

Example 3. The substitution $\tan x = t$ gives :

$$\int \frac{dx}{\sin x \cos x} = \int \frac{dx}{\tan x \cos^2 x} = \int \frac{dt}{t} = \ln |t| + C = \ln |\tan x| + C.$$

We shall now consider in detail a very important type of the primitive (1), *viz.* primitives of the kind

$$I_{m,n} = \int \sin^m x \cos^n x dx,$$

where m and n are integers. Evidently if the number $n = 2k + 1$ is odd, we have the conditions considered above in 1°, and the substitution $\sin x = t$ immediately rationalises the primitive (1); similarly if m is odd, the primitive is rationalised by the substitution $\cos x = t$ (case 2°); finally if both m and n are even or odd, we have the conditions described in 3° and the primitive is rationalised by the substitution $\tan x = t$. As we know from above, in every case the primitive can be rationalised by the substitution $\tan x/2 = t$. However, integration of the resulting function is often rather difficult. We must therefore look for other methods of integration; one such method which does not involve rationalisation is given below.

Integrating by parts when $n \neq -1$ we obtain :

$$\begin{aligned} I_{m,n} &= \int \sin^m x \cos^n x \, dx = \int \sin^{m-1} x (\cos^n x \sin x \, dx) = \\ &= -\frac{\sin^{m-1} x \cos^{n+1} x}{n+1} + \frac{m-1}{n+1} \int \sin^{m-2} x \cos^{n+2} x \, dx = \\ &= -\frac{\sin^{m-1} x \cos^{n+1} x}{n+1} + \frac{m-1}{n+1} I_{m-2, n+2}, \end{aligned} \quad (4)$$

and similarly when $m \neq -1$

$$\begin{aligned} I_{m,n} &= \frac{\sin^{m+1} x \cos^{n-1} x}{m+1} + \frac{n-1}{m+1} \int \sin^{m+2} x \cos^{n-2} x \, dx = \\ &= \frac{\sin^{m+1} x \cos^{n-1} x}{m+1} + \frac{n-1}{m+1} I_{m+2, n-2}. \end{aligned} \quad (5)$$

But

$$\begin{aligned} I_{m-2, n+2} &= \int \sin^{m-2} x \cos^{n+2} x \, dx = \\ &= \int \sin^{m-2} x \cos^n x (1 - \sin^2 x) \, dx = I_{m-2, n} - I_{m, n}, \end{aligned}$$

and similarly

$$I_{m+2, n-2} = I_{m, n-2} - I_{m, n};$$

consequently the formulae (4) and (5) give :

$$I_{m,n} = -\frac{\sin^{m-1} x \cos^{n+1} x}{n+1} + \frac{m-1}{n+1} [I_{m-2, n} - I_{m, n}], \quad (6)$$

$$I_{m,n} = \frac{\sin^{m+1} x \cos^{n-1} x}{m+1} + \frac{n-1}{m+1} [I_{m, n-2} - I_{m, n}], \quad (7)$$

and therefore we naturally also have

$$I_{m,n} = -\frac{\sin^{m-1} x \cos^{n+1} x}{m+n} + \frac{m-1}{m+n} I_{m-2, n}, \quad (8)$$

$$I_{m,n} = \frac{\sin^{m+1} x \cos^{n-1} x}{m+n} + \frac{n-1}{m+n} I_{m, n-2}. \quad (9)$$

We have thus obtained a pair of reduction formulae which enable us to diminish either of the two indices in the primitive $I_{m, n}$ (so long as the other index is not equal to -1 and $m+n \neq 0$) by two

units while the other index maintains its former value. If both numbers m and n are positive, then by successively applying both formulae we can evidently express the primitive $I_{m,n}$ in terms of any one of the primitives $I_{0,1}$, $I_{1,0}$, $I_{0,0}$, $I_{1,1}$ which can be evaluated directly.

The reduction formulae deduced above enable us to obtain results easily when one (or both) of the numbers m and n are negative. To prove this we at first note that the reduction formulae (8) and (9) remain valid when $m = -1$ and $n = -1$, which we did not include in the deduction of these formulae (this can be shown by a simple differentiation); the only exception is the case when $m + n = 0$. If we replace m by $m + 2$ in formula (8), we can express $I_{m,n}$ in terms of $I_{m+2,n}$, i.e. we can raise the first index by two units if $m + n + 2 \neq 0$; provided the same condition applies we can use formula (9) in order to raise the second index by two units. If $m + n + 2 = 0$ (with the exception of the case when $m = n = -1$), then either formula (6) or (7) enables us to find $I_{m,n}$; since we have $n \neq -1$ for $m + n + 2 = 0$, therefore by replacing m by $m + 2$ in formula (6) we readily obtain

$$I_{m,n} = \frac{\sin^{m+1} x \cos^{n+1} x}{m+1} + C.$$

In view of these facts we can evidently obtain integrable integrals by applying successively the formulae (8) and (9) when m and n are negative. Note that we have considered the only exception $I_{-1,-1}$ in the above example.

For problems to § 65 cf. Problem Book by B. P. Demidovich, Section III, Nos. 215-297, 299, 261, 265, 283, 304,

§ 66. Integration of differentials containing exponential functions

1. Let us consider the following primitive

$$I = \int e^{\alpha x} P(x) dx \quad (\alpha \neq 0), \quad (1)$$

where $P(x)$ is an arbitrary polynomial. Integrating by parts we obtain :

$$I = \frac{1}{\alpha} e^{\alpha x} P(x) - \frac{1}{\alpha} \int e^{\alpha x} P'(x) dx.$$

This formula replaces evaluation of the given primitive by evaluation of another simpler primitive, since the power of the polynomial $P'(x)$ is lower by unity than that of the polynomial $P(x)$. Repetition of this method gives :

$$I = \frac{e^{\alpha x}}{\alpha} \left\{ P(x) - \frac{1}{\alpha} P'(x) + \frac{1}{\alpha^2} P''(x) - \dots \right\} + C.$$

The resulting series is discontinuous so that the derivatives of the polynomial $P(x)$, from a certain order onwards, are identically zero. Hence all the primitives of the type (1) can be expressed by elementary functions.

We also note that the following primitives can be similarly found

$$I_1 = \int \sin \alpha x P(x) dx, \quad I_2 = \int \cos \alpha x P(x) dx \quad (\alpha \neq 0).$$

In fact, integration by parts gives :

$$I_1 = -\frac{\cos \alpha x}{\alpha} P(x) + \frac{1}{\alpha} \int \cos \alpha x P'(x) dx,$$

$$I_2 = \frac{\sin \alpha x}{\alpha} P(x) - \frac{1}{\alpha} \int \sin \alpha x P'(x) dx,$$

and successive application of these two formulae readily gives us final expressions for both primitives :

$$\begin{aligned} I_1 &= \frac{\sin \alpha x}{\alpha} \left\{ \frac{P'(x)}{\alpha} - \frac{P'''(x)}{\alpha^3} + \frac{P^{(5)}(x)}{\alpha^5} - \dots \right\} - \\ &\quad - \frac{\cos \alpha x}{\alpha} \left\{ P(x) - \frac{P''(x)}{\alpha^2} + \frac{P^{(4)}(x)}{\alpha^4} - \dots \right\} + C, \\ I_2 &= \frac{\sin \alpha x}{\alpha} \left\{ P(x) - \frac{P''(x)}{\alpha^2} + \frac{P^{(4)}(x)}{\alpha^4} - \dots \right\} + \\ &\quad + \frac{\cos \alpha x}{\alpha} \left\{ \frac{P'(x)}{\alpha} - \frac{P'''(x)}{\alpha^3} + \frac{P^{(5)}(x)}{\alpha^5} - \dots \right\} + C. \end{aligned}$$

Here also both series are automatically discontinuous.

2. Let us consider the primitives

$$K_1 = \int e^{\alpha x} \cos \beta x dx, \quad K_2 = \int e^{\alpha x} \sin \beta x dx \quad (\alpha \neq 0)$$

Integration by parts gives us :

$$K_1 = \frac{1}{\alpha} e^{\alpha x} \cos \beta x + \frac{\beta}{\alpha} \int e^{\alpha x} \sin \beta x \, d\alpha = \frac{1}{\alpha} e^{\alpha x} \cos \beta x + \frac{\beta}{\alpha} K_2,$$

$$K_2 = \frac{1}{\alpha} e^{\alpha x} \sin \beta x - \frac{\beta}{\alpha} K_1.$$

If we consider these two equations as a system of two equations with the unknowns K_1 and K_2 we obtain :

$$K_1 = \frac{e^{\alpha x} (\beta \sin \beta x + \alpha \cos \beta x)}{\alpha^2 + \beta^2} + C,$$

$$K_2 = \frac{e^{\alpha x} (\alpha \sin \beta x - \beta \cos \beta x)}{\alpha^2 + \beta^2} + C.$$

For problems to § 66 *cf.* Problem Book by B. P. Demidovich, Section III, Nos. 328, 330, 332, 333, 336.

CHAPTER XVIII

NUMERICAL INFINITE SERIES

§ 67. Fundamental concepts

Every branch of accurate nature study contains chapters describing the main concepts and laws of the given subject and other chapters devoted to the study of the subject ; these chapters must be learnt not so much in principle as in technique ; nevertheless, their methodical importance is often so great that study of the corresponding theory is often impossible without their systematic knowledge. Thus in the theory of heat we have chapters devoted to the measurement of temperature together with chapters dealing with theoretical problems *i.e.* practical methods for measuring temperatures.

The theory of infinite series plays a similar part in relation to the fundamental concepts and laws of mathematical analysis, *i.e.* it is a technical apparatus, an auxiliary tool ; also the numerous and varied applications of this apparatus are the basis of analysis and many applied sciences ; therefore the science of infinite series takes an outstanding place among contemporary mathematical methods and cannot be omitted from a comprehensive course of mathematical analysis.

The fundamental concept of the theory of infinite series is so elementary that everything connected with it could have been described much earlier in our course, for example, immediately after chapters devoted to the definition of limits and real numbers. We have postponed this subject for two reasons : we wanted to acquaint the reader as early as possible with the fundamental ideas differential and integral calculus ; we also wanted to connect the elementary concepts of infinite series with later chapters in which their further thorough development is described and the reader is only now able to understand this subject thoroughly.

The concept of an infinite series is very simple and completely expressed by summation of the following decreasing geometrical progression with which the reader should be well-acquainted from the high school stage :

$$a, ar, ar^2, \dots, ar^n, \dots, \quad (1)$$

where $0 < |r| < 1$ and a is an arbitrary real number. The sum of the first n terms of this progression

$$s_n = \sum_{k=0}^{n-1} ar^k = \frac{a - ar^n}{1 - r}$$

when $n \rightarrow \infty$ tends to the limit

$$\lim_{n \rightarrow \infty} s_n = \frac{a}{1 - r},$$

which is said to be sum of "all" terms of the given progression :

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1 - r}.$$

Hence in the case of a progression the summation of "all" terms of some infinite numerical series is carried out as follows : the sum s_n of the first n terms of the given series is constructed (it is evidently a function of n) and the behaviour of this sum for $n \rightarrow \infty$ is then investigated. If at the same time s_n tends to a definite limit s , we can naturally assume that s is the sum of "all" terms of this series.

A decreasing progression is, however, not the only series of this type. Thus, for example, the series

$$\frac{1}{1 \cdot 2}, \frac{1}{2 \cdot 3}, \frac{1}{3 \cdot 4}, \dots, \frac{1}{n(n+1)}, \dots \quad (2)$$

has exactly the same properties. In fact, since

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1} \quad (n = 1, 2, \dots),$$

therefore we have for the series (2) :

$$\begin{aligned} s_n = \sum_{k=1}^n \frac{1}{k(k+1)} &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \dots \\ &\dots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1 - \frac{1}{n+1}, \end{aligned}$$

and consequently

$$\lim_{n \rightarrow \infty} s_n = 1.$$

We have said above that $a / (1 - r)$ is the sum of "all" terms of the progression (1); we can now say for the same reason that sum of "all" terms of the series (2) is equal to 1 and we can write :

$$\sum_{k=1}^{\infty} k^{-\frac{1}{k+1}} = 1.$$

It is clear that every series cannot be summed in the manner described above. Thus for the series

$$1, 1, 1, \dots, 1, \dots$$

we have $s_n = n$; when $n \rightarrow \infty$, the sum s_n increases indefinitely and therefore cannot have a limit; the same phenomenon can take place when s_n is bounded; thus for the series

$$1, -1, 1, -1, \dots, 1, -1, \dots$$

s_n is equal to 0 or 1 depending on whether the number n is odd or even; when $n \rightarrow \infty$, the sum s_n remains bounded but does not tend to a limit.

We can now formulate general definitions. Let us assume that we are given an infinite sequence of real numbers

$$u_1, u_2, \dots, u_n, \dots \quad (3)$$

We assume that

$$s_n = \sum_{k=1}^n u_k \quad (n = 1, 2, \dots),$$

and call the sum s_n as *partial sum* of the series (3). If the limit

$$\lim_{n \rightarrow \infty} s_n = s$$

exists, we say that the series (3) is *convergent* and the number s is its sum; if, however, the partial sum s_n does not tend to a limit for $n \rightarrow \infty$, then the series (3) is *divergent* and has no sum.

The concept of sum of a convergent series and sum of "all" its terms appears to be self-evident, but one must be cautious; it must

be remembered that sum of an infinite series is not constructed like a finite sum, but an entirely new operation, that of limiting process, is involved in its construction; therefore the properties of finite sums cannot automatically be extended to sums of infinite series without first testing them; we shall see later that this extension is not possible in every case.

If the series (3) is convergent and its sum equal to s , we can write

$$s = \sum_{k=1}^{\infty} u_k = u_1 + u_2 + \dots + u_n + \dots;$$

if $s_n \rightarrow +\infty$ for $n \rightarrow \infty$, this is sometimes written as

$$\sum_{k=1}^{\infty} u_k = +\infty;$$

but it follows from the general definition that in this case the given series is *divergent* and has no sum.

Sometimes the following notation is used: $u_1 + u_2 + \dots +$

$+ u_n + \dots$ or $\sum_{k=1}^{\infty} u_k$ irrespective of whether the given series is convergent or divergent. For example, we say that the series

$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$ is convergent and the series $\sum_{k=1}^{\infty} (-1)^k$ divergent (cf.

above examples).

If the series (3) is convergent, the difference $r_n = s - s_n$ between its sum and partial sum is called *remainder* of the given series; it follows from the definition of convergence that

$$r_n \rightarrow 0 \quad (n \rightarrow \infty);$$

the remainder r_n of a convergent series is an infinitely small quantity for $n \rightarrow \infty$ (a divergent series has, of course, neither a remainder nor a sum). Since

$$s = u_1 + u_2 + \dots + u_n + u_{n+1} + \dots, \quad s_n = u_1 + u_2 + \dots + u_n,$$

therefore we naturally expect that

$$r_n = s - s_n = u_{n+1} + \dots + u_{n+k} + \dots = \sum_{k=1}^{\infty} u_{n+k} = \sum_{l=n+1}^{\infty} u_l.$$

This equation which is obvious for finite sums cannot be extended to infinite series without proof. However, the proof happens to be very simple; if we assume that

$$\sum_{k=1}^r u_{n+k} = \sigma_r \quad (r = 1, 2, \dots),$$

then evidently $\sigma_r = s_{n+r} - s_n$; and owing to the fact that $s_{n+r} \rightarrow s$ for $r \rightarrow \infty$ and n is constant, σ_r has a limit equal to $s - s_n = r_n$ for $r \rightarrow \infty$;

but by definition the limit of σ_r is the sum of the series $\sum_{k=1}^{\infty} u_{n+k}$;

therefore this series is convergent and its sum is r_n , which was to be proved. Thus if the series (3) converges, we have for every $n \geq 1$

$$s = s_n + r_n,$$

where

$$s_n = u_1 + u_2 + \dots + u_n, \quad r_n = u_{n+1} + u_{n+2} + \dots + u_{n+k} + \dots.$$

It follows from our definition of convergence of the series (3) that this convergence is equivalent to the condition that the sequence of partial sums

$$s_1, s_2, \dots, s_n, \dots \tag{4}$$

should tend to a definite limit s which in this case is said to be the sum of the series. Hence the sum and convergence of the series (3) depend on existence and magnitude of the limit of the sequence (4). Thus infinite series can be expressed in terms of sequences and their limits. But it can be readily seen that the relationship between series and sequences is mutual. Let us assume that we are given the sequence (4) of arbitrary real numbers s_n and that

$$u_1 = s_1, \quad u_n = s_n - s_{n-1} \quad (n > 1);$$

we then evidently have

$$u_1 + u_2 + \dots + u_n = s_n \quad (n = 1, 2, \dots),$$

and existence and magnitude of the limit of the sequence (4) depends entirely on convergence and sum of the series (3).

This elementary relationship between sequences and infinite series can often be conveniently used to apply propositions proved for one of these subjects to the other without additional proof. In § 19 (theorem 2) we have proved the following necessary and sufficient condition for existence of a limit for the sequence (4) : in order that the sequence (4) should have a limit it is necessary and sufficient that the following condition be satisfied : no matter how small $\varepsilon > 0$, we should have $|s_{n+p} - s_n| < \varepsilon$ for every sufficiently large n and $p > 0$. But if the numbers s_n are the partial sums of the series (3), we have

$$s_{n+p} - s_n = u_{n+1} + u_{n+2} + \dots + u_{n+p} = \sum_{k=1}^p u_{n+k},$$

and, on the other hand, existence of the limit of the sequence (4) is equivalent to the convergence of the series (3); we thus arrive at the following necessary and sufficient condition for convergence of series.

Theorem 1. *In order that the series (2) should be convergent it is necessary and sufficient that the following condition be satisfied : no matter how small $\varepsilon > 0$, $|u_{n+1} + u_{n+2} + \dots + u_{n+p}| < \varepsilon$ for every sufficiently large n and for every $p > 0$.*

This condition can be picturesquely expressed as follows : the absolute value of every sufficiently far removed "part" of the series (irrespective of the length of this "part", i.e. of the number of terms of the series it contains) should be as small as we please. Thus when $p = 1$, it follows from this condition that $|u_n| < \varepsilon$ for every convergent series (3), provided n is sufficiently large; in other words, we have :

Corollary : *If the series (2) is convergent, then $u_n \rightarrow 0$ for $n \rightarrow \infty$.*

In the examples of divergent series which we have so far considered u_n does not tend to zero for $n \rightarrow \infty$; therefore the question may arise whether the condition $u_n \rightarrow 0$ ($n \rightarrow \infty$) which, as we have just shown, is necessary for convergence of the series (3) is also sufficient for this purpose. It can be readily shown that this not so. In fact, let the series (3) be constructed as follows : At first $u_1 = 1$ is taken ; then two terms (u_2 and u_3) each of which is equal to $\frac{1}{2}$ follow ;

then follow three more terms, each of which is equal to $\frac{1}{3}$; this is continued ad infinitum. It is then obvious that on one hand $u_n \rightarrow 0$ for $n \rightarrow \infty$. On the other hand, the "part" of the series consisting of the terms $1/k$ contains, by its construction, k terms and is therefore equal to unity. And since k can be as large as we please, the "parts" equal to unity will occur in the constructed series as far removed as we please. The condition of theorem 1 is not satisfied and the series (3) does not converge.

A classical example of this kind is provided by the very instructive "harmonic" series

$$1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} + \dots; \quad (5)$$

the condition $u_n \rightarrow 0$ ($n \rightarrow \infty$) is satisfied, but the "part" of the series

$$\sum_{n=2^{k+1}}^{2^{k+1}-1} \frac{1}{n}$$

contains $2^{k+1} - 2^k = 2^k$ terms, neither of which is smaller than the preceding term which is equal to $1/2^{k+1}$; therefore this "part" is greater than

$$2^k \cdot \frac{1}{2^{k+1}} = \frac{1}{2},$$

and since we can select each "part" as far removed as we please (the number k being arbitrarily large), the condition of theorem 1 is again not satisfied and the series (5) is divergent.

Infinite series, like other general concepts, need more details for their full development; their full meaning can only be interpreted by considering more or less specialised classes of series which are characterised by individual properties. In this paragraph we are studying the concept of infinite series in its most general form and there remains something besides to say.

Let us assume that we are given two series :

$$\begin{aligned} u_1 + u_2 + \dots + u_n + \dots, \\ v_1 + v_2 + \dots + v_n + \dots \end{aligned}$$

In that case the series

$$(u_1 + v_1) + (u_2 + v_2) + \dots + (u_n + v_n) + \dots \quad (6)$$

can be regarded as the result of "term-by-term" addition of the two given series *i.e.* an addition where each term of the first series is added to the corresponding term of the second series. Let us assume that both series are convergent. Let us denote their respective sums by s and σ and partial sums by s_n and σ_n so that

$$s_n \rightarrow s, \quad \sigma_n \rightarrow \sigma \quad (n \rightarrow \infty).$$

In this case the sum of the first n terms of the series (6) will evidently be equal to $s_n + \sigma_n$ and tend to $s + \sigma$ for $n \rightarrow \infty$; thus term by-term addition of two convergent series results in another convergent series and the sum of the new series is equal to the sum of the two given series. This rule evidently remains valid (and is proved by the same method) if the series are subtracted term-by-term.

Finally, no radical change in our arguments will take place if, instead of two series, we take an arbitrary finite number of convergent series and construct an arbitrary algebraic sum with an arbitrary combination of signs (which must naturally be the same for all terms). The series resulting from this algebraic term-by-term addition of the given series will always be convergent and its sum will be equal to the result of the algebraic addition (with the same combination of signs) of the sums of the given series. We thus arrive at the following proposition :

Theorem 2. *Let the series*

$$\sum_{k=1}^{\infty} u_{1,k}, \quad \sum_{k=1}^{\infty} u_{2,k}, \quad \dots, \quad \sum_{k=1}^{\infty} u_{m,k}$$

be convergent and their sums be respectively equal to $s_1, s_2, \dots, \dots, s_m$. Then the series

$$\sum_{k=1}^{\infty} (u_{1,k} \pm u_{2,k} \pm \dots \pm u_{m,k})$$

(where the same combination of signs is taken for each term) will also be convergent and its sum will be equal to

$$s_1 \pm s_2 \pm \dots \pm s_m.$$

It follows from this theorem that by changing a *finite number* of terms of a convergent series we cannot affect its convergence (although we are in general changing its sum); in other words, the following result holds:

Corollary. *If in the series*

$$u_1 + u_2 + \dots + u_n + u_{n+1} + \dots \quad (7)$$

and

$$v_1 + v_2 + \dots + v_n + v_{n+1} + \dots \quad (8)$$

we have

$$u_{n+1} = v_{n+1}, u_{n+2} = v_{n+2}, \dots, u_{n+k} = v_{n+k}, \dots$$

for $n \geq 0$ and if one of these series is convergent, then the other series is also convergent.

In order to prove, it is sufficient to note that the series (8) is obtained by term-by-term addition of the series (7) and the (convergent) series

$$(v_1 - u_1) + (v_2 - u_2) + \dots + (v_n - u_n) + 0 + 0 + \dots$$

It obviously follows from the above corollary that if one of the series (7) and (8) is divergent, then the other series is also divergent.

Another general property of numerical series is established by the following theorem :

Theorem 3. *If the series*

$$u_1 + u_2 + \dots + u_n + \dots$$

is convergent and its sum is equal to s and if a is an arbitrary constant, then the series

$$au_1 + au_2 + \dots + au_n + \dots$$

is also convergent and its sum is equal to as .

In order to prove this theorem it is sufficient to note that, denoting by s_n and σ_n the respective sums of the first n terms of the above two series, we should have $\sigma_n = as_n$ for any n .

For exercises to § 67 *cf.* Problem Book by B. P. Demidovich, Section V, Nos. 1-3, 5, 11-12, 14, 15, 21, 23, 24.

§ 68. Series with constant signs

We have already remarked above that in order to develop the meaning of infinite series thoroughly we must study specific classes of series characterised by special properties which make them important and accessible for study. The history of the development of the

science of infinite series has revealed that the most important class of this kind are series whose all terms have the same sign. Therefore we shall at first study series with "constant signs." We shall always assume in such cases that all terms are positive (or, more accurately, non-negative, since, in general, it is useful to assume existence of terms equal to zero). It is obvious that series with negative terms (or, more accurately, non-positive terms) will possess analogous properties owing to symmetry.

If all terms of the series

$$u_1 + u_2 + \dots + u_n + \dots \quad (1)$$

are non-negative and if we assume, as above, that

$$\sum_{k=1}^n u_k = s_n \quad (n = 1, 2, \dots),$$

then we shall evidently have $s_{n+1} \geq s_n$ for all $n \geq 1$, i.e. the partial sums s_n form a non-decreasing sequence. However, in this case there are only two possibilities as $n \rightarrow \infty$: the sum s_n may increase indefinitely as $s_n \rightarrow +\infty$, or it can remain bounded; in the latter case, as we know, it should tend to a definite limit s ; but the relation $s_n \rightarrow s$ ($n \rightarrow \infty$) implies that the series (1) is convergent and that its sum is equal to s . Hence *in order that a series with constant signs should be convergent it is necessary and sufficient that its partial sums should be bounded*. In the general case when the terms of the series have different signs this condition evidently also remains necessary for the series to converge; however, it is no longer the sufficient condition as can be seen from the divergent series considered in § 67:

$$1 + (-1) + 1 + (-1) + \dots,$$

whose partial sums are only equal to 1 or 0 and which are therefore bounded.

The condition established above gives us a very valuable criterion for convergence of series with constant signs; convergence and applications of many series occurring in analysis can be established by the direct or indirect use of this criterion. At the same time the above criterion is also valuable from a theoretical point of view; because of it the theory of series with constant signs becomes very clear and accessible and can be developed much further than the theory of series with variable signs, to which the above condition

cannot be applied. Its importance can be easily understood. We have found in our initial definition that in order to determine convergence of the given series we have to study the quantity s_n as a function of n so as to find whether it tends to a definite limit for $n \rightarrow \infty$; the expression for s_n in terms of n is often rather complicated and does not directly reveal the limiting behaviour of this quantity as $n \rightarrow \infty$. On the other hand, it is often sufficient to assess roughly the quantity s_n in order to confirm that it remains bounded for $n \rightarrow \infty$; if the series has constant signs, we can directly deduce that s_n has a limit and the given series converges.

Let us consider an example. It is required to find whether the series

$$\frac{1}{2+1} + \frac{1}{2^2+1} + \frac{1}{2^3+1} + \dots + \frac{1}{2^n+1} + \dots$$

converges. The expression obtained for the sum of its first n terms is complicated and does not conclude as regards limiting behaviour of this quantity. However, since

$$\frac{1}{2^k+1} < \frac{1}{2^k} \quad (k = 1, 2, \dots, n),$$

therefore

$$s_n < \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^n} = 1 - \frac{1}{2^n} < 1$$

for every $n \geq 1$. Hence the quantity s_n is bounded and the given series converges.

The method used in the above example often enables us to determine convergence of concrete series. In its general form it can be formulated as follows:

Theorem 1. *We are given two series with non-negative terms :*

$$u_1 + u_2 + \dots + u_n + \dots, \quad (u)$$

$$v_1 + v_2 + \dots + v_n + \dots, \quad (v)$$

if a positive number c and a natural number n_0 exist in this case such that

$$u_n \leq cv_n$$

for every $n \geq n_0$, then convergence of the series (v) implies convergence of the series (u) and, conversely, divergence of the series (u) implies divergence of the series (v).

This theorem is sometimes known as the "*principle of comparison of series*".

Proof. It follows from corollary 2 § 67 that we can evidently assume without loss of generality that the inequality $u_n \leq cv_n$ is valid for all n ; let us denote by s_n and σ_n the respective partial sums of the series (u) and (v) ; we evidently have $s_n \leq c \sigma_n$ ($n = 1, 2, \dots$); if the series (v) converges, then the sums σ_n are bounded and therefore the sum s_n is also bounded which in its turn implies convergence of the series (u) . Theorem 1 is thus proved.

The principle of comparison of series established above can be applied not only to the study of definite series but also to the deduction of many convenient tests of convergence which find frequent applications. We shall establish several such tests in the sequel.*)

Test 1 (Cauchy). *If a positive number $r < 1$ exists such that*

$$\sqrt[n]{u_n} \leq r$$

for all sufficiently large n , then the series (u) is convergent; if, however, values of n exist which can be as large as we please, for which

$$\sqrt[n]{u_n} \geq 1,$$

then the series (u) is divergent.

Proof. In the first case we have for all sufficiently large n

$$u_n \leq r^n,$$

and, according to theorem 1, convergence of the series (u) follows from convergence of the progression $r + r^2 + \dots + r^n + \dots$. In the second case we have $u_n \geq 1$ for an infinite number of values of n and divergence of the series (u) follows from the corollary of theorem 1 § 67.

In particular, if the limit

$$\lim_{n \rightarrow \infty} \sqrt[n]{u_n}$$

exists, the above test enables us to establish the following simple result :

Corollary. *If the series (u) has $\lim_{n \rightarrow \infty} \sqrt[n]{u_n} = l$, then the series (u) is convergent for $l < 1$ and divergent for $l > 1$.*

*) Up to the end of § 68 we have considered series which we assume to have non-negative terms.

In fact, let $l < 1$ and $\varepsilon > 0$ be so small that $l + \varepsilon < 1$; it follows from $\sqrt[n]{u_n} \rightarrow l$ ($n \rightarrow \infty$) that $\sqrt[n]{u_n} < l + \varepsilon$ for all sufficiently large n and therefore, according to the test 1, the series (u) is convergent. If, however, $l > 1$, then $\sqrt[n]{u_n} > 1$ for all sufficiently large n and, according to the test 1, the series (u) diverges.

In case $\lim_{n \rightarrow \infty} \sqrt[n]{u_n} = 1$, the above corollary does not enable us to draw any conclusions with regards to convergence of the series (u) . The fact that the series (u) may in this case be divergent can be seen from the simple series

$$1 + 1 + \dots + 1 + \dots$$

However, we shall soon see that the series (u) can also be convergent in this case.

Test 2. (D' Alembert). *If a positive number $r < 1$ exists such that*

$$\frac{u_{n+1}}{u_n} \leq r,$$

for all sufficiently large n , then the series (u) converges; if, however, for all sufficiently large n

$$\frac{u_{n+1}}{u_n} \geq 1,$$

then the series (u) diverges.

Proof. In the first case we have for all sufficiently large values of n

$$u_{n+1} \leq u_n r,$$

$$u_{n+2} \leq u_{n+1} r \leq u_n r^2,$$

$$u_{n+3} \leq u_{n+2} r \leq u_n r^3,$$

and generally

$$u_{n+k} \leq u_n r^k \quad (k = 1, 2, \dots);$$

hence, according to theorem 1, convergence of the series (u) follows from convergence of the progression

$$u_n r + u_n r^2 + \dots + u_n r^k + \dots$$

In the second case the terms of the series (u) form, from a certain term onwards, a non-decreasing sequence of positive numbers; the relation $u_n \rightarrow 0$ ($n \rightarrow \infty$) is therefore impossible and the series (u) diverges in view of the corollary of theorem 1 § 67.

Like test 1, test 2 gives the following result :

Corollary. *If the series (u) has $\lim_{n \rightarrow \infty} u_{n+1}/u_n = l$, then the series (u) converges for $l < 1$ and diverges for $l > 1$.*

The proof is analogous to that of the corollary of test 1 and we leave it to the reader. As before, if $l = 1$, we cannot conclude that the series (u) is convergent.

Example 1. Consider the series

$$\frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{n!} + \dots;$$

here

$$\frac{u_{n+1}}{u_n} = \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \frac{1}{n+1} \rightarrow 0 \quad (n \rightarrow \infty);$$

$l = 0$; the series converges; its sum is equal to $e - 1$, as can be easily calculated on the basis of examples considered in § 39.

For further exercises cf. Problem Book by B. P. Demidovich, Section V, Nos. 28-32.

It is obvious from the proofs of the above two tests that they can serve for comparison of the given series with a geometrical progression. These two tests can, however, only be successfully applied to series whose terms decrease more rapidly than the terms of a geometrical progression. Such series must be regarded as “roughly” convergent—their terms decrease very rapidly and therefore the remainder r_n also tends rapidly to zero as n increases, i.e. the partial sum rapidly tends to its limit s . We say that such series “converge rapidly”; the more rapidly a series converges, the more convenient it is for practical calculations; if, for example, s_4 already gives us s with a degree of accuracy sufficient for the given problem, then all we have to do is to add four terms of our series; if the series converges slowly, we may have to calculate, say, s_{100} in order to obtain the same degree of accuracy and this will be a much more complicated technical operation. For this reason it is sometimes said that a very slowly convergent series may appear to be “practically divergent”; in spite of the fact that s_n can be as close as we please to s when n is sufficiently large, we have to take a very large

value of n so as to obtain the required degree of accuracy and we cannot evaluate the sum in practice.

Our two tests cannot be applied to series which converge "slower" than an arbitrary geometrical progression. This means that for such series the limit l mentioned in both corollaries usually appears to be equal to unity. More accurate and sensitive tests must be found for these series and many efforts were made in this direction, since many extensive classes of series of great practical importance belong to these "slowly" convergent series.

Let us begin by considering a more important class of series of the kind

$$\frac{1}{1^s} + \frac{1}{2^s} + \dots + \frac{1}{n^s} + \dots, \quad (2)$$

where s is an arbitrary constant real number. If $s \leq 0$, the series (2) evidently diverges; we are therefore only interested in positive values of s . If $s = 1$, we have a harmonic series whose divergence has been established in § 67. And since we have for $s < 1$

$$\frac{1}{n^s} \geq \frac{1}{n} \quad (n = 1, 2, \dots),$$

it therefore follows from the principle of comparison of series (theorem 1) that divergence of the harmonic series implies divergence of the series (2) for every $s < 1$. We must therefore only consider the values $s > 1$.

We will now show that the series (2) converges for every $s > 1$. Let $k > 1$ be an arbitrary natural number. Since we evidently have $x^{-s} \geq k^{-s}$ for $0 < x \leq k$, therefore

$$\int_{k-1}^k x^{-s} dx \geq \int_{k-1}^k k^{-s} ds = \frac{1}{k^s};$$

assuming in this inequality that $k = 2, 3, \dots, n$ and adding the inequalities thus obtained we have :

$$\sum_{k=2}^n \frac{1}{k^s} \leq \int_1^n x^{-s} dx = \frac{1}{s-1} - \frac{1}{(s-1)n^{s-1}} < \frac{1}{s-1},$$

since $s > 1$. This shows that the set of partial sums of the series (2) is bounded for $s > 1$ which, as we know, is sufficient for the series (2) to converge.

The method by which we have proved convergence of the series (2) for $s > 1$ is very often used ; we shall give it a general basis in § 107 (theorem 5). At present we only note that this method enables us not only to establish convergence of the series (2) but also to give a convenient assessment to its remainder. In fact, on the basis of what has been proved above we have for $k > 1$

$$\int_k^{k+1} x^{-s} dx \leq \frac{1}{k^s} \leq \int_{k-1}^k x^{-s} dx.$$

Summing this inequality with respect to k from n to $n+r$ we obtain :

$$\int_n^{n+r+1} x^{-s} dx \leq \sum_{k=n}^{n+1} \frac{1}{k^s} \leq \int_{n-1}^{n+r} x^{-s} dx,$$

or, evaluating the integrals

$$\begin{aligned} \frac{1}{(s-1)n^{s-1}} - \frac{1}{(s-1)(n+r+1)^{s-1}} &\leq \sum_{k=n}^{n+r} \frac{1}{k^s} \leq \\ &\leq \frac{1}{(s-1)(n-1)^{s-1}} - \frac{1}{(s-1)(n+r)^{s-1}}; \end{aligned}$$

when $\rightarrow \infty$, these inequalities give in the limit :

$$\frac{1}{(s-1)n^{s-1}} \leq \sum_{k=n}^{\infty} \frac{1}{k^s} \leq \frac{1}{(s-1)(n-1)^{s-1}}.$$

Example 2. When $s = 3$, we have :

$$\frac{1}{2n^2} \leq \frac{1}{n^3} + \frac{1}{(n+1)^3} + \frac{1}{(n+2)^3} + \dots \leq \frac{1}{2(n-1)^2}.$$

Convergence of the series (2) could not be proved by means of either of the above tests. In particular, the limits

$$\lim_{n \rightarrow \infty} \sqrt[n]{u_n}, \quad \lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n},$$

mentioned in the above corollaries, are equal to unity for the series (2) for all values of s . In fact, assuming

$$\sqrt[n]{n^{-s}} = c_n \quad (n = 1, 2, \dots)$$

we have :

$$\ln c_n = -s \frac{\ln n}{n},$$

and consequently (cf. example 5 § 37)

$$\ln c_n \rightarrow 0, \quad c_n \rightarrow 1$$

for $n \rightarrow \infty$. On the other hand

$$\frac{\frac{1}{(n+1)^s}}{\frac{1}{n^s}} = \left(1 + \frac{1}{n}\right)^{-s} \rightarrow 1 \quad (n \rightarrow \infty)$$

irrespective of s .

In the same way in which the above two tests were based on comparison with a geometrical progression, other more sensitive tests of convergence can be constructed on the basis of a comparison with a series of the type (2). We shall now prove one such test.

Test 3 (Raabe). *If a number $r > 1$ exists such that*

$$n \left(\frac{u_n}{u_{n+1}} - 1 \right) \geq r \tag{3}$$

for all sufficiently large values of n , then the series

$$u_1 + u_2 + \dots + u_n + \dots \tag{u}$$

converges ; if, however, for all sufficiently large values of n

$$n \left(\frac{u_n}{u_{n+1}} - 1 \right) \leq 1, \tag{4}$$

then the series (u) diverges.

Proof 1. In the first case let us denote by r' an arbitrary number between 1 and r ($1 < r' < r$). Evidently the quantity

$$\frac{\left(1 + \frac{1}{n}\right)^{r'} - 1}{\frac{1}{n}}$$

has as its limit the derivative of the function $(1+x)^{r'}$ at $x=0$ for $n \rightarrow \infty$, i.e. the number r' ; since $r' < r$, we have for all sufficiently large values of n

$$\frac{\left(1 + \frac{1}{n}\right)^{r'} - 1}{\frac{1}{n}} < r,$$

hence

$$\left(1 + \frac{1}{n}\right)^{r'} < 1 + \frac{r}{n}.$$

But it follows from (3) that in this case

$$\frac{u_n}{u_{n+1}} > 1 + \frac{r}{n} > \left(1 + \frac{1}{n}\right)^{r'} = \left(\frac{n+1}{n}\right)^{r'},$$

or

$$n^{r'} u_n > (n+1)^{r'} u_{n+1};$$

this shows that for sufficiently large values of n the product $n^{r'} u_n$ decreases in the transition from n to $n+1$ and therefore remains bounded for $n \rightarrow \infty$; in other words, a number $c > 0$ exists such that

$$n^{r'} u_n < c \quad (n = 1, 2, \dots),$$

or

$$u_n < \frac{c}{n^{r'}} \quad (n = 1, 2, \dots).$$

Since $r' > 1$, the series $\sum_{n=1}^{\infty} \frac{1}{n^{r'}}$, as we have just proved, is convergent; it therefore follows from the principle of comparison of series that the series (u) is also convergent.

2. In the second case it follows from (4) that

$$\frac{u_n}{u_{n+1}} \leq \frac{n+1}{n},$$

or

$$n u_n \leq (n+1) u_{n+1};$$

therefore provided n is sufficiently large ($n \geq n_0$), the product $n u_n$ does not decrease in the transition from n to $n+1$; hence if we assume that $n_0 u_{n_0} = c$, we shall have $n u_n \geq c$ for $n \geq n_0$ and consequently

$$u_n \geq \frac{c}{n},$$

and since a harmonic series diverges, it follows from the principle of comparison of series that the series (u) also diverges. We have thus proved test 3.

Corollary. *If the series (3) has the following limit*

$$\lim_{n \rightarrow \infty} n \left(\frac{u_n}{u_{n+1}} - 1 \right) = l,$$

then the series (u) converges for $l > 1$ and diverges for $l < 1$.

We can again leave the proof to the reader. When $l = 1$, the given series can either converge or diverge.

Example 1. Let us assume that for $n \geq 2$

$$u_n = a^{-\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n-1}\right)},$$

where a is a constant positive number. We readily obtain

$$\frac{u_n}{u_{n+1}} = a^{\frac{1}{n}},$$

hence

$$n \left(\frac{u_n}{u_{n+1}} - 1 \right) = \frac{a^{\frac{1}{n}} - 1}{\frac{1}{n}}.$$

The limit of this expression for $n \rightarrow \infty$ is evidently equal to the derivative of the function a^x with respect to x at $x = 0$, i.e. $\ln a$. It follows from test 3 that the series (u) therefore converges for $a > e$ and diverges for $a < e$. The problem needs further study for $a = e$.

We could readily construct simple examples of series for which the test 3 is too rough. We have just met one such series ($a = e$ in the last example). The reader can show for himself that this test tells us nothing about convergence of the series of the type

$$\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^s},$$

where s is a constant positive number. We could find a more sensitive test to throw some light on this problem. In chapter XXV

we shall study a test of completely different type based on integral calculus which will enable us to operate freely with series of this and other more complicated types.

For exercises cf. Problem Book by B.P. Demidovich, Section V, Nos. 42 and 45.

§ 69. Series with variable signs

We shall now begin to study series with arbitrary signs.

Among them we find the so-called *alternating* series whose terms have alternately positive and negative signs so that, for example, all terms with odd indices are positive. Such a series can be written in the form

$$u_1 - u_2 + u_3 - u_4 + \dots + u_{2k-1} - u_{2k} + \dots, \quad (1)$$

where all u_n are obviously positive numbers; alternating series occur very frequently and have many different applications; they are also interesting theoretically; their convergence can often be established by means of a very simple test which we shall now prove.

Theorem 1 (Leibnitz). *If the inequality $u_{n+1} \leq u_n$ holds for any $n \geq 1$ and if $\lim_{n \rightarrow \infty} u_n = 0$, then the alternating series (1) is convergent.*

Hence for alternating series the tendency of u_n to zero for $n \rightarrow \infty$ together with the monotonic decrease of the absolute value of terms guarantees convergence of the series (in general this is, of course, not so; let us remind you, for example, of a harmonic series which satisfies both these conditions).

Proof. Let us denote, as usual, the partial sums of the series (1) by s_n . We have for every $k \geq 2$

$$s_{2k} - s_{2k-2} = u_{2k-1} - u_{2k} \geq 0;$$

hence the sequence

$$s_2, s_4, s_6, \dots, s_{2k}, \dots \quad (2)$$

is non-decreasing. But on the other hand

$$s_{2k} = u_1 - (u_2 - u_3) - (u_4 - u_5) - \dots - (u_{2k-2} - u_{2k-1}) - u_{2k},$$

hence

$$s_{2k} \leq u_1 \quad (k = 1, 2, \dots),$$

so that all minuends on the right-hand side are non-negative; therefore the non-decreasing sequence (2) is bounded from above and has a limit

$$\lim_{k \rightarrow \infty} s_{2k} = s.$$

In order to prove that the partial sums s with odd indices tend to the same limit, it is sufficient to note that, according to proof,

$$s_{2k+1} = s_{2k} + u_{2k+1}$$

and that $s_{2k} \rightarrow s$ for $k \rightarrow \infty$; it is also given that $u_{2k+1} \rightarrow 0$; therefore

$$s_{2k+1} \rightarrow s \quad (k \rightarrow \infty),$$

which proves theorem 1.

A typical simple series which often occurs in applications is the following series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots + \frac{1}{2k-1} - \frac{1}{2k} + \dots, \quad (3)$$

which according to theorem 1 is convergent. If we replace all terms of this series by their absolute values, we obtain a divergent (harmonic) series. This shows that convergence of the series (3) depends not so much on the rate of decrease of its terms (their absolute values) as on the alternative distribution of their signs.

The series (3) also shows that it is possible for a given series to converge while the series composed of the absolute values of its terms diverges. It is most important in the theory of all series with variable signs that the converse of the situation described above should not occur: if the series composed of the absolute values of terms of the given series converges, then the given series will always be convergent. Let

$$u_1 + u_2 + \dots + u_n + \dots \quad (u)$$

be a series composed of terms with arbitrary signs. In that case the following theorem holds:

Theorem 2. *If the following series is convergent*

$$|u| + |u| + \dots + |u_n| + \dots, \quad (|u|),$$

then the series (u) is also convergent.

Proof. Let n and k be two arbitrary natural numbers. According to theorem 1 § 67 convergence of the series $(|u|)$ implies that the sum

$$|u_{n+1}| + |u_{n+2}| + \dots + |u_{n+k}|$$

is as small as we please when n is sufficiently large and k arbitrary; owing to the fact that

$$|u_{n+1} + u_{n+2} + \dots + u_{n+k}| \leq |u_{n+1}| + |u_{n+2}| + \dots + |u_{n+k}|,$$

the same also holds for the quantity $|u_{n+1} + u_{n+2} + \dots + u_{n+k}|$; according to theorem 1 § 67 this also implies convergence of the series (u) . Theorem 2 is therefore proved.

We can thus see that convergent series (u) with alternating signs can be divided into two categories: series for which the series $(|u|)$ converges and the other series for which it diverges. Series of the first type are said to be *absolutely convergent* and those of the second type *conditionally convergent* (the reason for this terminology will soon be obvious). The difference in properties of these two types of convergent series is very remarkable and fundamentally important in analysis and many applications. This difference can be basically characterised by the fact that absolutely convergent series possess almost all properties of finite sums, since all operations with such series are carried out according to the same rules as with finite sums; on the other hand, many simple properties of finite series which are very important in applications do not apply to conditionally convergent series and therefore practical application of these series becomes rather restricted.

It is evident that convergent series with constant signs always converge absolutely so that the difficulties mentioned above do not arise. All propositions which we are going to establish in the next paragraph for absolutely convergent series also hold for all series with constant signs.

In conclusion to this paragraph we shall give one more test for convergence of series with variable signs. Let $\alpha_1, \alpha_2, \dots, \alpha_n \dots$ and $\beta_1, \beta_2, \dots, \beta_n \dots$ be two sequences of real numbers which possess the following properties: 1) the number α_n are positive and decrease monotonically ($\alpha_{n+1} < \alpha_n$) and $\lim_{n \rightarrow \infty} \alpha_n = 0$; 2) a positive number c exists such that $|\sigma_n| = |\beta_1 + \beta_2 + \dots + \beta_n| < c$ for every $n \geq 1$, i.e. the series of numbers β_n has bounded partial sums. Let

us assume that $u_n = \alpha_n \beta_n$ ($n = 1, 2, \dots$) and show that the given series (u) converges.

For this purpose let us apply the general test for convergence (theorem 1 §67) according to which it is sufficient to show that for every n and for an arbitrary $p > 0$

$$|u_{n+1} + u_{n+2} + \dots + u_{n+p}| < \varepsilon.$$

We have

$$\begin{aligned} \rho(n, p) &= u_{n+1} + u_{n+2} + \dots + u_{n+p} = \\ &= \alpha_{n+1} \beta_{n+1} + \alpha_{n+2} \beta_{n+2} + \dots + \alpha_{n+p} \beta_{n+p}, \end{aligned}$$

or assuming that

$$\beta_1 + \beta_2 + \dots + \beta_k = \sigma_k \quad (k = 1, 2, \dots),$$

$$\begin{aligned} \rho(n, p) &= \alpha_{n+1} (\sigma_{n+1} - \sigma_n) + \alpha_{n+2} (\sigma_{n+2} - \sigma_{n+1}) + \dots \\ &\dots + \alpha_{n+p} (\sigma_{n+p} - \sigma_{n+p-1}) = -\sigma_n \alpha_{n+1} + \sigma_{n+1} (\alpha_{n+1} - \alpha_{n+2}) + \\ &+ \sigma_{n+2} (\alpha_{n+2} - \alpha_{n+3}) + \dots + \sigma_{n+p-1} (\alpha_{n+p-1} - \alpha_{n+p}) + \sigma_{n+p} \alpha_{n+p}; \end{aligned}$$

let us choose n so large that $\alpha_{n+1} < \varepsilon/2c$; it then follows from the inequalities $|\sigma_k| < c$ and $\alpha_{k+1} \leq \alpha_k$ which hold for every k that the last equation gives us, regardless of p :

$$|\rho(n, p)| \leq c\alpha_{n+1} + c(\alpha_{n+1} - \alpha_{n+p}) + c\alpha_{n+p} = 2c\alpha_{n+1} < \varepsilon,$$

which proves our proposition. We thus arrive at the following test:

Theorem 3 (Dirichlet). *Let $\alpha_n \rightarrow 0$ ($n \rightarrow \infty$) and let the following inequalities hold for every $n \geq 1$:*

$$\alpha_{n+1} \leq \alpha_n, \quad |\beta_1 + \beta_2 + \dots + \beta_n| < c,$$

where c is a constant. In this case the series

$$\alpha_1 \beta_1 + \alpha_2 \beta_2 + \dots + \alpha_n \beta_n + \dots$$

will also converge.

By choosing, say, $\beta_n = (-1)^{n-1}$ we exactly obtain theorem 1 as can readily be seen; hence the latter is a particular case of theorem 3.

Example. We have for every k

$$\sin\left(k + \frac{1}{2}\right)x - \sin\left(k - \frac{1}{2}x\right) = 2 \sin \frac{x}{2} \cos kx.$$

Summing this relation with respect to k from 1 to n we obtain

$$\sin\left(n + \frac{1}{2}\right)x - \sin \frac{1}{2}x = 2 \sin \frac{x}{2} \sum_{k=1}^n \cos kx,$$

and assuming that $\sin \frac{1}{2}x \neq 0$

$$\sum_{k=1}^n \cos kx = \frac{\sin\left(n + \frac{1}{2}\right)x - \sin \frac{1}{2}x}{2 \sin \frac{x}{2}},$$

and therefore for every $n \geq 1$

$$\left| \sum_{k=1}^n \cos kx \right| < \frac{1}{\left| \sin \frac{x}{2} \right|}.$$

Assuming that $\alpha_n = 1/n$, $\beta_n = \cos nx$ we can conclude, according to theorem 3, that the series

$$\sum_{n=1}^{\infty} \frac{\cos nx}{n} \quad (4)$$

converges, provided x is not a multiple of 2π ; when $x = \pi$, the series (4) is transformed into the series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n},$$

whose convergence, has been proved above for $x = 2\pi$; the series (4) then becomes a harmonic series.

For further exercises to § 69 *cf.* Problem Book by B.P. Demidovich, Section V, Nos. 74-77, 85-86, 89, 96.

§ 70. Operations with series

1. One of the most important properties of finite sums is *commutativity*, *i.e.* the sum is independent of the order of terms; we are therefore naturally interested in finding whether this property also holds for infinite series, *i.e.* whether the sum of an infinite convergent series remains unchanged when its terms are arbitrarily commutated

and whether the series remains convergent. We shall now learn that for absolutely convergent series and for conditionally convergent series this problem is solved in a directly opposite sense.

Theorem 1. *If the series*

$$u_1 + u_2 + \dots + u_n + \dots \quad (u)$$

is absolutely convergent and its sum is equal to s , then the series

$$v_1 + v_2 + \dots + v_n + \dots, \quad (v)$$

obtained from it as a result of an arbitrary commutativity of the numbers u_n , is also absolutely convergent and its sum is also s .

Proof. Let us assume that

$$\rho_n = \sum_{k=n+1}^{\infty} |u_k|,$$

so that $\rho_n \rightarrow 0$ ($n \rightarrow \infty$). Let ε be an arbitrary positive number and let n be such that $\rho_n < \varepsilon$. The numbers u_1, u_2, \dots, u_n of the series (u) coincide with some definite numbers $v_{i_1}, v_{i_2}, \dots, v_{i_n}$ of the series (v). Let m_n be the greatest of the indices i_1, i_2, \dots, i_n . In this case the sum

$$\sigma_m = \sum_{k=1}^m v_k$$

for every $m \geq m_n$ evidently contains among its terms all the numbers u_k ($1 \leq k \leq n$); it may also contain in addition some numbers u_k with indices $k > n$. Therefore assuming that

$$\sum_{k=1}^n u_k = s_n \quad (n = 1, 2, \dots),$$

we have :

$$\sigma_m = s_n + q$$

where q is the sum of several numbers u_k with indices $k > n$ so that

$$|q| \leq \sum_{k=n+1}^{\infty} |u_k| = \rho_n < \varepsilon.$$

Therefore

$$\begin{aligned} |\sigma_m - s| &= |s_n - s + q| \leq |s_n - s| + |q| = \\ &= \left| \sum_{k=n+1}^{\infty} u_k \right| + |q| \leq \sum_{k=n+1}^{\infty} |u_k| + |q| < 2\varepsilon, \end{aligned}$$

the only condition being that m is sufficiently large. Hence the series (v) is convergent and its sum is equal to s . Absolute convergence of this series is almost self-evident; in fact, the sum

$$\sum_{k=1}^n |v_k|$$

is no other than the sum of n terms of a convergent series with non-negative coefficients $|u_1| + |u_2| + \dots + |u_n| + \dots$ and therefore it does not exceed the sum of this series for any n ; it therefore remains bounded for $n \rightarrow \infty$ and this implies convergence of the series

$$\sum_{k=1}^{\infty} |v_k|, \text{ i.e. absolute convergence of the series } (v).$$

We shall now turn to conditionally convergent series and at first prove one auxiliary proposition for these series, which is also of general interest.

Lemma. *If the series (u) is conditionally convergent, then all its positive terms form a divergent series (u^+) and, similarly, all its negative terms form a divergent series (u^-) .*

Proof. We shall denote respectively by s_n^+ and s_n^- the sums of those terms of the series (u^+) and (u^-) which have the partial sums s_n of the series (u) so that $s_n^+ + s_n^- = s_n$. Since the series (u) is convergent, the sum s_n tends to a definite limit for $n \rightarrow \infty$; hence the equation $s_n = s_n^+ + s_n^-$ shows that if either of the sums s_n^+ or s_n^- has a limit for $n \rightarrow \infty$, then the other sum must also tend to a limit but in this case the difference $s_n^+ - s_n^-$ must also tend to a definite limit which is evidently equal to

$$|u_1| + |u_2| + \dots + |u_n|;$$

hence the series (u) is absolutely convergent, which contradicts the conditions of the theorem. Therefore neither s_n^+ nor s_n^- can tend to a limit for $n \rightarrow \infty$; i.e. the series (u^+) and (u^-) diverge, which was to be proved.

Theorem 2. *If the series (u) is conditionally convergent, then an appropriate commutativity of its terms can make it divergent or convergent, and in the latter case its sum can be made equal to a preassigned sum s .*

Proof 1. In order to obtain a divergent series we shall place the terms of the series (u) as follows: At first we take a certain number of positive terms of the series (u) so that their sum should exceed unity (this is possible as a result of the lemma which we have just proved). Thereafter we place the first negative term; we then take enough positive terms for their sum to exceed unity and we place thereafter the second negative term; it follows from our lemma that this process can be continued *ad infinitum* and it is evident that every term of the series (u) will sooner or later be included in the new series. Our series will, by construction, contain "parts" > 1 which can be as far removed as we please; therefore according to theorem 1 § 67 the series will be divergent.

2. In order to obtain a convergent series with an arbitrarily preassigned sum s we shall place the terms of the series (u) as follows: We assume, say, that $s \geq 0$. Therefore, at first we take positive terms of the series (u) (in the same order as they appear in this series until their sum does not exceed s ; this will take place sooner or later as a result of the lemma proved above; as soon as the sum obtained exceeds s we begin to add negative terms of the series (u) (again in their natural order); we continue to do so until their sum becomes less than s ; this will again take place sooner or later as a result of the same lemma. As soon as this happens we again begin to add positive terms of the series (u) , and so on. The resulting series

$$v_1 + v_2 + \dots + v_n + \dots \quad (v)$$

will, in fact, include all terms of the series (u) but their position will differ. Let us assume that

$$\sum_{k=1}^n v_k = \sigma_n \quad (n = 1, 2, \dots).$$

Let $\epsilon > 0$ be as small as we please. Since $v_n \rightarrow 0$ for $n \rightarrow \infty$, we can find an m such that $|v_n| < \epsilon$ for $n \geq m$. Let us consider an arbitrary sum σ_n ($n \geq m$); if σ_n and σ_{n-1} lie on opposite sides of s , then

$$|\sigma_n - s| < |\sigma_n - \sigma_{n-1}| = |v_n| < \epsilon.$$

If, however, σ_n and σ_{n-1} lie on the same side of s , then according to our construction σ_n lies closer to s than σ_{n-1} . Hence in all cases σ_n lies closer to s than the distance ε or it lies closer than the preceding sum σ_{n-1} . This evidently implies that from a certain number onwards all the sums σ_n lie closer to s than the distance ε ; and since ε is arbitrarily small, $\sigma_n \rightarrow s (n \rightarrow \infty)$, which was to be proved.

We can therefore see that with regards to commutativity of terms a conditionally convergent series represents, as it were, a raw, amorphous mass which can be transformed, by suitably applying the above operation, into either a divergent series or a convergent series or a convergent series with a preassigned sum.

Note. The problem about the influence of commutativity of terms of the series which we have considered above arises, as is almost self-evident, only when this operation embraces an infinite number of terms of the series. In fact, if only terms with indices not exceeding m can be commutated, then all partial sums of the series, beginning with s_m , remain unchanged; if the initial series is convergent, then the series obtained after this commutative operation will also be convergent and its sum will be the same.

2. Another important property of finite sums is their distributivity: in order to multiply two sums we must multiply each term of one sum by all the terms of the other and add the product so obtained. It is therefore important to know whether this distributive law also applies to infinite series. This problem can also be considered from another point of view: we have seen § 67 that two (or more) convergent series can always be added or subtracted term-by-term. We are now naturally interested in finding whether it is permissible to multiply these series term-by-term.

Theorem 3. *If the series*

$$u_1 + u_2 + \dots + u_n + \dots = s \quad (u)$$

and

$$v_1 + v_2 + \dots + v_n + \dots = \sigma \quad (v)$$

are absolutely convergent, then the series composed of all products of the form $u_i v_k$ ($i, k = 1, 2, \dots$) with suffixes appearing in any order is also absolutely convergent and its sum is equal to $s\sigma$.

Proof. Let us denote by $w_1, w_2, \dots, w_n, \dots$ products of the form $u_i v_k$ ($i, k = 1, 2, \dots$) numbered in any order and consider the series

$$|w_1| + |w_2| + \dots + |w_n| + \dots \quad (|w|)$$

Let S_n ($n = 1, 2, \dots$) be the partial sums of this series. The sum S_n consists of terms of the form $|u_i v_k|$. Among the suffixes i and k of terms composing the series S_n we can find the greatest term; let us denote it by m ; if then we multiply term-by-term the finite sums

$$A_m = |u_1| + |u_2| + \dots + |u_m|, \quad B_m = |v_1| + |v_2| + \dots + |v_m|,$$

the terms of this product will evidently contain all the terms $|u_i v_k|$ of the sum S_n . Hence

$$S_n \leq A_m B_m.$$

But the series (u) and (v) are absolutely convergent and therefore the sums A_m and B_m are bounded; the last inequality shows that the partial sums S_n of the series $(|w|)$ are bounded and, consequently, this series is convergent.

We have to prove that the sum of the series

$$w_1 + w_2 + \dots + w_n + \dots \quad (w)$$

(whose convergence follows, of course, from convergence of the series $(|w|)$ is equal to $s\sigma$. With this in mind we note that since the series (w) is absolutely convergent, therefore, in order to find its sum we can place its terms (*i.e.* the products $u_i v_k$) in any order (theorem 1). Let us place them as follows: at first we take the term $u_1 v_1$ (only) in which the greatest suffix is unity; we then take all those terms whose greatest suffixes are equal to 2 (there will be three such numbers: $u_1 v_2, u_2 v_2, u_2 v_1$; after that we take terms with greatest suffix equal to 3 (there will be five such terms: $u_1 v_3, u_2 v_3, u_3 v_3, u_3 v_2, u_3 v_1$), and so on. If we take the partial sum of this series (w) ending with a group of terms with the greatest suffix m , then its partial sum evidently includes all products of the form $u_i v_k$ ($1 \leq i \leq m, 1 \leq k \leq m$), *i.e.* it will be equal to $s_m \sigma_m$, where

$$s_m = u_1 + u_2 + \dots + u_m, \quad \sigma_m = v_1 + v_2 + \dots + v_m;$$

and since we have $s_m \rightarrow s$ and $\sigma_m \rightarrow \sigma$ for $m \rightarrow \infty$, the selected partial sum of the series (w) tends to $s\sigma$ for $m \rightarrow \infty$. And since the series (w) is convergent, the limit of this partial sum should coincide with the sum of the series (w) which is therefore equal to $s\sigma$. This proves theorem 3.

A more accurate analysis which we cannot consider here shows that in order to multiply term-by-term the series (u) and (v) it is sufficient to assume that either of these series is absolutely convergent (and, of course, we must also assume that the other series is at least

conditionally convergent). In that case we can no longer place the products $u_i v_k$ arbitrarily, but we must place them in a quite definite order.

When both series are conditionally convergent, term by term multiplication is almost impossible. Hence, in general, only absolutely convergent series possess all distributive properties of finite sums.

For exercises *cf.* Problem Book by B.P. Demidovich, Section V, Nos. 116, 119.

§ 71. Infinite products

As we know, addition can embrace an arbitrary number of terms; similarly multiplication can also embrace as many factors as we please. In the case of addition by allowing the number of terms to increase indefinitely and using the principle of limiting process we have arrived at the concept of a sum of an infinite series. Owing to the fact that addition is in many respects similar to multiplication we can expect to arrive at a new useful concept, when the number of factors increases indefinitely, by applying the principle of limiting process.

Let z_1, z_2, \dots, z_n be an arbitrary sequence of real numbers. Let us assume that

$$z_1 z_2 \dots z_n = \prod_{k=1}^n z_k = \pi_n \quad (n = 1, 2, \dots)$$

and call the numbers π_n the “partial products” of the given sequence. If the limit

$$\lim_{n \rightarrow \infty} \pi_n = \pi$$

exists, then, as with infinite series, we can naturally consider this number π to be the product of “all” the numbers z_n and write

$$\pi = \prod_{k=1}^{\infty} z_k = z_1 z_2 \dots z_n \dots$$

Let us assume that all the numbers z_n are positive. In that case

$\log \pi_n = \sum_{k=1}^n \log z_k$; if the limit (1) exists and if $\pi \neq 0$, then it follows

from $\pi_n \rightarrow \pi$ and from continuity of the logarithmic function that $\log \pi_n \rightarrow \log \pi$, *i.e.*

$$s_n = \sum_{k=1}^n \log z_k \rightarrow \log \pi \quad (n \rightarrow \infty).$$

This shows that existence of the *positive limit* (1) inevitably implies convergence of the series

$$\log z_1 + \log z_2 + \dots + \log z_n + \dots, \quad (2)$$

and the sum of this series is equal to $\log \pi$. Conversely, if the series (2) is convergent, then the quantity

$$s_n = \sum_{k=1}^n \log z_k = \log (z_1 z_2 \dots z_n) = \log \pi_n$$

tends to a definite limit for $n \rightarrow \infty$; hence the partial product π_n also has a definite limit (this limit has a logarithm and is therefore positive). We thus arrive at the conclusion that *in order that the nonzero limit (1) should exist (for positive z_n) it is necessary and sufficient that the series (2) should be convergent*. This enables us to foresee that the case $\pi = 0$ will deserve special attention in the theory of infinite products.

Let us now assume that the numbers z_k have arbitrary signs (we only assume that there are no numbers equal to zero among them: if, for example, $z_k = 0$, then evidently $\pi_n = 0$ for all $n \geq k$ and the limiting behaviour of π_n becomes trivial. We evidently have :

$$z_n = \frac{\pi_n}{\pi_{n-1}} \quad (n = 2, 3, \dots).$$

Let the limit (1) exist; in that case $\pi_n \rightarrow \pi$ and $\pi_{n-1} \rightarrow \pi$ for $n \rightarrow \infty$ and therefore if $\pi \neq 0$

$$z_n \rightarrow \frac{\pi}{\pi} = 1 \quad (n \rightarrow \infty);$$

in the same way as the n -th term of a convergent series should tend to zero for $n \rightarrow \infty$, so the n -th factor of an infinite product should tend to unity for $n \rightarrow \infty$, provided a nonzero limit π exists. We can see that here the case $\pi = 0$ occupies a special place; it can readily be

shown by an example that for $\pi = 0$ our deduction will, in general, be no longer true ; for this purpose we can choose, say,

$$z_n = \frac{1}{2} \quad (n = 1, 2, \dots) :$$

we have $\pi_n = 1 / 2^n \rightarrow 0$ ($n \rightarrow \infty$), while z_n is always equal to $1/2$ and therefore does not tend to unity.

We have thus confirmed for the second time that products with a nonzero limit (1) show a more or less close analogy with convergent series. This is confirmed by many other facts in the further development of theory. It is therefore useful to accept the following definition.

The infinite product

$$z_1 z_2 \dots z_n \dots = \prod_{n=1}^{\infty} z_n$$

is said to be convergent if the nonzero limit (1) exists ; if, however, this limit either does not exist or, although existing, is equal to zero, then the infinite product is said to be divergent.

If the infinite product converges, the limit π is said to be its *value* or *quantity* ; thus the value of an infinite product is always expressed by a nonzero number, a divergent product has no value (meaning).

We have seen that the n -th term of a convergent series tends to zero as $n \rightarrow \infty$. This shows that in every convergent product (3) the numbers z_n , from a certain n onwards, are always positive ; the product can therefore only contain a finite number of negative factors ; if we change the sign of each factor, the product as a whole will either change its sign or the sign may remain the same, *i.e.* it will only undergo a very trivial change. Hence without loss of generality we can assume (and we will do so in future) that all the numbers z_n are positive. Moreover, owing to the fact that for convergent products $z_n \rightarrow 1$ ($n \rightarrow \infty$), it is frequently convenient to assume that $z_n = 1 + u_n$ and write the infinite product in the form

$$\prod_{n=1}^{\infty} (1 + u_n) ; \tag{4}$$

here we always have $-1 < u_n < +\infty$, and in the case of a convergent product $u_n \rightarrow 0$ ($n \rightarrow \infty$).

One of the main aims of the theory of infinite products is to establish tests which would enable us to recognise whether the given infinite product is convergent or divergent. As we know, convergence of the product (4) is equivalent to convergence of the series

$$\sum_{n=1}^{\infty} \log (1 + u_n) ; \quad (5)$$

this connection enables us to base the tests for convergence of products on the known tests for convergence of series as established in § 67 ; it enables us to confirm the condition that for every $\varepsilon > 0$ and for every sufficiently large n and $k > 0$ we should have :

$$\left| \sum_{i=n+1}^{n+k} \log (1 + u_i) \right| < \varepsilon. \quad (6)$$

this is the necessary and sufficient condition for convergence of the series (5) and therefore also of the product (4). Since

$$\sum_{i=n+1}^{n+k} \log (1 + u_i) = \log \left\{ \prod_{i=n+1}^{n+k} (1 + u_i) \right\},$$

the inequality (6) can be replaced by the inequality *)

$$\left| \prod_{i=n+1}^{n+k} (1 + u_i) - 1 \right| < \eta,$$

where η , like ε , is a positive number which can be as small as we please. Hence in order that an infinite product should be convergent it is necessary and sufficient that any sufficiently far removed (as far as we please) "part" of the series should be as close to unity as we please. Hence in this respect we have a complete analogy between products and series. Let us draw attention to the fact that this analogy only holds if we accept the above definition (*i.e.* the condition that products with $\lim \pi_n = \pi = 0$ are divergent).

The established test is of great theoretical importance like the corresponding test for series ; however, it can only be applied in isola-

*) This number is as close to unity as we please if and only if its logarithm is as small as we please.

ted cases of concrete products. In order to obtain tests of greater practical value we must, as in the theory of series, go beyond the general concept and consider definite classes of infinite products. We must at first naturally ask which products can be regarded as analogous to series with constant signs and which is analogous to series with variable signs.

The terms of a convergent series tend to zero as their suffix increases; if the series has constant signs, this implies that all its terms are either positive or negative; in other words, all its terms lie on the same side of the limiting value 0. In the case of a convergent infinite product the limiting value of all terms is unity; we must therefore regard those infinite products as analogous to series with constant signs in which all factors are either greater or less than unity. If the product is represented in the form (4), this means that the numbers u_n are either all positive or all negative. For infinite products of this kind we have a very simple and practically convenient criterion of convergence :

Theorem 1. *If all the numbers u_n have the same sign, then in order that the product (4) should be convergent it is necessary and sufficient that the following series should be convergent :*

$$u_1 + u_2 + \dots + u_n + \dots \quad (7)$$

Proof. We can at first assume that $u_n \rightarrow 0$ for $n \rightarrow \infty$. In fact, if this is not so, then, as we know, the product (4) and the series (7) are divergent and the statement of theorem 3 is proved. Let us now consider both possibilities separately.

1. Let $u_n \geq 0$ ($n = 1, 2, \dots$). Owing to the fact that $x \rightarrow 0$

$$e^x = 1 + x + o(x), \quad (8)$$

therefore for a sufficiently large n

$$\frac{1}{e^2} u_n \leq 1 + u_n \leq e^{2u_n}.$$

We can assume that these inequalities are satisfied for *all* n , since rejection of a finite number of terms cannot affect convergence of the series or the product. But assuming that

$$\sum_{k=1}^n u_k = s_n, \quad \prod_{k=1}^n (1 + u_k) = \pi_n,$$

we have :

$$e^{\frac{1}{2}s_n} \leq \pi_n \leq e^{2s_n} \quad (n = 1, 2, \dots). \quad (9)$$

If the series (7) is convergent, then the sum s_n remains bounded for $n \rightarrow \infty$; it then follows from the second of the inequalities (9) that π_n also remains bounded; and since $\pi_{n+1} \geq \pi_n$ ($n \geq 1$), the product (4) is convergent; conversely, if the product (4) is convergent, then π_n remains bounded for $n \rightarrow \infty$; it then follows from the first of the inequalities (9) that s_n also remains bounded and therefore the series (7) is convergent.

2. $u_n \leq 0$ ($n = 1, 2, \dots$). The same relation (8) gives us for a sufficiently large n

$$e^{2u_n} \leq 1 + u_n \leq e^{\frac{1}{2}u_n},$$

and we can again assume that these inequalities are satisfied for *all* values of n , which leads to

$$e^{2s_n} \leq \pi_n \leq e^{\frac{1}{2}s_n} \quad (n = 1, 2, \dots). \quad (10)$$

If the series (7) is convergent, then s_n is convergent for $n \rightarrow \infty$ (this time $s_n \leq 0$ and the fact that s_n is bounded implies existence of a positive number A such that $s_n > -A$ for any n); it follows from the first of the inequalities (10) that $\pi_n \geq e^{-2A}$ ($n = 1, 2, \dots$) and owing to the fact that now $\pi_{n+1} \leq \pi_n$ ($n \geq 1$), therefore π_n tends to a definite *positive* limit $\pi \geq e^{-2A}$ for $n \rightarrow \infty$, i.e. the product (4) converges. Conversely, if the product (4) converges, then $\pi_n \rightarrow \pi > 0$ ($n \rightarrow \infty$) and, evidently, $\pi_n \geq \pi$ ($n = 1, 2, \dots$); the second of the inequalities (10) therefore gives:

$$e^{\frac{1}{2}s_n} \geq \pi, \quad s_n \geq 2 \ln \pi \quad (n = 1, 2, \dots),$$

i.e. the sum s_n remains bounded from below for $n \rightarrow \infty$; hence the series (7) converges.

Examples. It follows from divergence of the harmonic series that for $n \rightarrow \infty$

$$\left(1 + \frac{1}{1}\right)\left(1 + \frac{1}{2}\right) \dots \left(1 + \frac{1}{n}\right) \rightarrow +\infty,$$

$$\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right) \dots \left(1 - \frac{1}{n}\right) \rightarrow 0.$$

It follows from convergence of the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

that both products

$$\left(1 + \frac{1}{1^2}\right)\left(1 + \frac{1}{2^2}\right) \dots \left(1 + \frac{1}{n^2}\right) \text{ and } \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right) \dots \left(1 - \frac{1}{n^2}\right),$$

tend to positive limits for $n \rightarrow \infty$.

For further exercises *cf.* Problem Book by B.P. Demidovich, Section V, Nos. 403, 413, 419, 420, 425.

CHAPTER XIX

INFINITE SERIES OF FUNCTIONS

§ 72. Region of Convergence of a Series of Functions

Let $u_1(x)$, $u_2(x)$, ..., $u_n(x)$ be a sequence of functions with independent variable x defined in an interval (a, b) . If we write the infinite series

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots, \quad (1)$$

then for each value x_0 of the variable x in (a, b) this series becomes a numerical series

$$u_1(x_0) + u_2(x_0) + \dots + u_n(x_0) + \dots,$$

which can be convergent or divergent. A series of the type (1) is said to be an *infinite series of functions* defined in the interval (a, b) . These series are an important tool in mathematical analysis, and the whole theory of numerical series whose elements were described in Chapter XVIII can be regarded as an introduction to the theory of series of functions which we shall now study.

Let us at first consider how the concept of convergence can be applied to series of functions. We have already said above that for every numerical value of the variable x in the interval (a, b) the series (1) becomes a numerical series so that in view of Chapter XVIII the expression (1) can be regarded as describing not a single series but a whole family of numerical series. Some of these series will be convergent and some divergent. It is therefore obvious that we cannot give one answer to the question whether the series (1) converges or diverges; such a question should not be put for series of functions. Instead the following question should be asked: *For what values of x in the interval (a, b) is the series (1) convergent and for what values divergent?* Hence convergence of a series of functions is a *local* concept: it applies

at some points of the given interval (a, b) and does not apply at other points. Only when the series (1) converges (or diverges) at *every* point in the interval (a, b) , it can be said that the series converges (or diverges) in that interval.

A point x in the interval (a, b) at which the series (1) converges is said to be the *point of convergence* of this series; similarly the point at which the series (1) diverges is said to be the *point of divergence* of this series. Hence in relation to every series defined in the interval (a, b) the points of this interval are divided into two sets of points: the set of points of convergence and the set of points of divergence of the series (1). The first set is known as *region of convergence* and the second as *region of divergence* of the given series. In some cases either set may be empty.

Cases are known in the theory of series of functions when regions of convergence and divergence have very complicated structures; this may be so even when the series are composed of simple elementary functions; it comes about, for example, with the trigonometrical series which we shall study in chapter XXI. Here we shall only consider one simple case.

The series

$$1 + x + x^2 + \dots + x^n + \dots,$$

all terms of which are defined along the whole number line $(-\infty < x < +\infty)$ is a geometrical progression for every value of x ; the region of convergence of this series is evidently the open interval $-1 < x < 1$; the region of divergence is defined by the inequality $|x| \geq 1$.

Sums of the form

$$s_n(x) = \sum_{k=1}^n u_k(x),$$

in analogy with numerical series, are called *partial sums* of the series (1). If the series (1) converges at the point x , then the following limit exists: $\lim_{n \rightarrow \infty} s_n(x) = s(x)$. The functions $s_n(x)$ are defined at

every point of the interval (a, b) , but the function $s(x)$, which is said to be *sum* of the series (1), is only defined at the points of convergence of this series. The function $r_n(x) = s(x) - s_n(x)$ is called *remainder* of the given series. It is evident that, regardless of the value of n , the

function $r_n(x)$ is only defined within the region of convergence of the series (1). At every point x of this region we have:

$$\lim_{n \rightarrow \infty} r_n(x) = 0.$$

§ 73. Uniform convergence

We have already noted that convergence of a series of functions

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots$$

is local in character. When we say that a given series converges in an interval (a, b) we mean that it converges at every single point in this interval, but this does not deprive the concept of convergence of its local character. We can, however, introduce another concept of convergence of a series of functions in a given interval, which does not imply its convergence at individual points and which has a "total" and not a "local" character. This concept is of fundamental importance in the theory of series of functions and its applications, and we shall now consider it in detail.

Let the series (1) whose partial sum is denoted by $s_n(x)$ be convergent at every point in the interval (a, b) and let its sum be equal to $s(x)$; the remainder of the series $r_n(x) = s(x) - s_n(x)$ tends to zero as $n \rightarrow \infty$ at any point x in the interval (a, b) . In detail this implies as follows: for every $\varepsilon > 0$ and for any $x (a \leq x \leq b)$ a natural number n_0 can be found such that for every $n \geq n_0$

$$|r_n(x)| < \varepsilon.$$

This natural number n_0 , i.e. the "spot" from which the inequality (2) is satisfied, evidently depends not only on ε but also on the selected point x in the interval (a, b) . Thus different values of x produce different numerical series (1); therefore, in general, the "spot" from which $|r_n(x)|$ always remains less than ε will be different for different series. Is it possible to choose n_0 such that for every $n \geq n_0$ the inequality (2) should be satisfied for all values of x in (a, b) ? If a finite number of values of x exists, this problem would be simple; each value of x would then correspond to a definite value of n_0 so that there would only be a finite number of values of n_0 ; having selected the greatest values of n_0 we would evidently obtain the "spot" from which the inequality (2) would be satisfied for all values of x (of which there is a finite number). However, the interval (a, b) contains not a finite but an infinite number of values of x , each of

which has a corresponding n_0 so that we have an infinite number of values of n_0 ; an infinite set of natural numbers does not always contain the greatest number; (we must therefore take into account the possibility that the number n_0 , from which "spot" onwards the inequality (2) would be satisfied at every point of the interval (a, b) , does not exist). We can also see the reason for this: for every point x on the given line the "spot", from which we always have $|r_n(x)| < \varepsilon$, will appear sooner or later; for some points this will happen sooner and for others later; for some values of x the series (1) converges more rapidly and for other values of x more slowly; we can say that convergence of the series at some points "lags behind", as it were, from its convergence at other points, *i.e.* although the series converges for all values of x ($a \leq x \leq b$), its convergence is "non-uniform" and takes place more rapidly for some values of x and more slowly for other values.

Therefore the following definition appears useful.

The series (1) is uniformly convergent in the interval (a, b) if, no matter what the number $\varepsilon > 0$ be, a natural number n_0 can be found such that the inequality (2) holds for every $n \geq n_0$ and for every x ($a \leq x \leq b$).

This new concept of convergence of a series of functions is no longer local in character, *i.e.* it cannot entirely be reduced to convergence of the series at individual points and it essentially takes into account the comparative rate of convergence at different points. We must at first consider *non-uniformly* convergent series. Can the series (1) which converges at every point in the interval (a, b) not be non-uniformly convergent in that interval? Let us recall that a similar problem has been considered in § 23; having defined, the local continuity of a function at every point of a given interval we proceed to define the concept of *uniform* continuity which is more restrictive and no longer local in character; however, it appeared later (theorem 5§ 23) that every continuous function is also uniformly continuous (in a closed interval), *i.e.* the new concept is no more restrictive than the initial concept. We have a similar situation in this case if every series (1) which is convergent at every point in the interval (a, b) is also uniformly convergent in that interval. However, we shall now show that this is not so.

Let us assume that

$$u_1(x) = x, u_n(x) = x^n - x^{n-1} \quad (n > 1)$$

and consider the series (1) in the interval $(0, 1)$. We have:

$$s_n(x) = x + (x^2 - x) + \dots + (x^n - x^{n-1}) = x^n.$$

This gives us for $0 \leq x < 1$

$$\lim_{n \rightarrow \infty} s_n(x) = \lim_{n \rightarrow \infty} x^n = 0,$$

while $s_n(1) = 1$ ($n = 1, 2, \dots$) for $x = 1$, and therefore

$$\lim_{n \rightarrow \infty} s_n(1) = 1.$$

Hence assuming that

$$s(x) = \begin{cases} 0 & (0 \leq x < 1), \\ 1 & (x = 1), \end{cases}$$

we have :

$$s_n(x) \rightarrow s(x) \quad (0 \leq x \leq 1);$$

in other words, the series (1) converges at every point of the interval $(0, 1)$ and its sum is equal to $s(x)$.

We shall now show that this convergence is non-uniform. The point

$$x_n = \frac{1}{\sqrt[n]{2}}$$

evidently lies in the interval $(0, 1)$ for every natural n ; but

$$s_n(x_n) = x_n^n = \frac{1}{2}, \quad s(x_n) = 0,$$

and therefore

$$r_n(x_n) = s(x_n) - s_n(x_n) = -\frac{1}{2}, \quad |r_n(x_n)| = \frac{1}{2}.$$

If $\varepsilon < 1/2$ (for example $\varepsilon = 1/4$), then no matter how large the value of n be, a point x_n can be found in the interval $(0, 1)$, at which

$$|r_n(x_n)| > \varepsilon;$$

the inequality (2) cannot be satisfied at *every* point of the interval $(0, 1)$, no matter how large the value of n be; this means that our series is non-uniformly convergent in the interval $(0, 1)$.

It is very important to imagine this phenomenon visually. Fig. 47 represents the graphs of the functions $y = s_n(x)$ for $n = 1$, $n = 2$ and for a very large value of n . We can see that no matter which point x ($0 \leq x < 1$) we choose, the value of $s_n(x)$ decreases and tends to zero as n increases; it can be seen from the graph that

this value is negligibly small for large n . However, no matter how large the values of n we choose for values of x close to unity (for example, for $x = 1/n/2$), $s_n(x)$ will still be far away from its limit (i.e. from zero): points can be found on the curve $y = s_n(x)$ such that their ordinates are still far away from zero; and if we continue to increase the number n repeatedly, such points will still appear (they will only change their positions by moving further to the right or left). The “lagging” of convergence which we have mentioned above can thus be vividly imagined.

The reader will readily appreciate that no matter how small $\varepsilon > 0$ be, the given series will converge uniformly in the interval $(0, 1 - \varepsilon)$. Hence only the behaviour of terms of the series in the immediate neighbourhood of the point unity inhibits uniform convergence of this series in the interval $(0, 1)$.

We can thus see that a series of functions can converge non-uniformly in the given interval. This means that the concept of

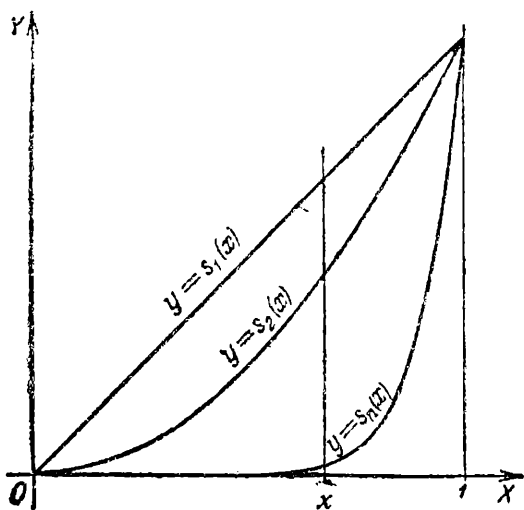


Fig. 47.

convergence of a series, which we have introduced above, is more restrictive than the concept of local convergence of a series in an interval. In the next few paragraphs we shall consider several general problems in which the concept of uniform convergence is of fundamental importance. At present, however, we shall only consider tests which will enable us to establish uniform convergence of a series in a given interval.

To begin with, there holds the following necessary and sufficient condition for uniform convergence (which is therefore theoretically very valuable) which is analogous to theorem 1 § 67 for numerical series.

Theorem 1. *In order that the series (1) should be uniformly convergent in the interval (a, b) , it is necessary and sufficient that the following condition should be satisfied: no matter how small $\varepsilon > 0$ be, the following inequality hold will for all sufficiently large values of n :*

$$\left| \sum_{k=n+1}^{n+p} u_k(x) \right| = |u_{n+1}(x) + u_{n+2}(x) + \dots + u_{n+p}(x)| < \varepsilon, \quad (3)$$

irrespective of the natural number p and the point x in the interval (a, b) .

Proof. 1. If the series (1) converges uniformly in the interval (a, b) , then for $n \geq n_0$ and for every natural p

$$|r_n(x)| < \frac{\varepsilon}{2}, \quad |r_{n+p}(x)| < \frac{\varepsilon}{2} \quad (a \leq x \leq b),$$

and therefore

$$|r_{n+p}(x) - r_n(x)| = \left| \sum_{k=n+1}^{n+p} u_k(x) \right| < \varepsilon \quad (a \leq x \leq b),$$

which proves the necessity of our condition.

$$2. \quad \text{Owing to the fact that } r_n(x) = \lim_{p \rightarrow \infty} \sum_{k=n+1}^{n+p} u_k(x), \text{ therefore in}$$

the case when the inequality (3) holds for every natural number p and at every point x in the interval (a, b) we have :

$$|r_n(x)| \leq \varepsilon \quad (a \leq x \leq b), \quad (4)$$

and it therefore follows from the condition of theorem 1 that the inequality (4) will be satisfied for every $\varepsilon > 0$ provided n is sufficiently large; this means that the series (1) is uniformly convergent in the interval (a, b) ; hence we have proved the sufficiency of the condition.

In concrete cases uniform convergence is frequently established by means of the following simple and very convenient sufficient condition.

Theorem 2 (Weierstrass' test). *If the following numerical convergent series with positive terms exists :*

$$a_1 + a_2 + \dots + a_n + \dots, \quad (5)$$

so that

$$|u_n(x)| \leq a_n \quad (n = 1, 2, \dots; \quad a \leq x \leq b), \quad (6)$$

then the series (1) is uniformly convergent in the interval (a, b) .

Proof. No matter how small $\varepsilon > 0$ be, it follows from theorem 1 § 67 that the following inequality would hold for all sufficiently large values of n

$$a_{n+1} + a_{n+2} + \dots + a_{n+p} < \varepsilon,$$

regardless of the natural number p . But it then follows from the inequalities (6) that

$$\left| \sum_{k=1}^p u_{n+k}(x) \right| \leq \sum_{k=1}^p |u_{n+k}(x)| < \varepsilon \quad (a \leq x \leq b),$$

provided n is sufficiently large and p is an arbitrary natural number. We therefore conclude from theorem 1 that the series (1) is uniformly convergent in (a, b) ; theorem 2 is thus proved.

For exercises to § 73 cf. Problem Book by B. P. Demidovich, Section V, Nos. 242, 243, 245, 246, 261, 262, 264, 268, 279, 285, 286.

§ 74. Continuity of sum of a series of functions

In the study of numerical series we were considering as to which properties of finite sums also hold for infinite series, *i.e.* to what extent we can treat series in the same way as finite sums. We are naturally also interested in the same problem with regard to the theory of series of functions.

We know that sum of a finite series of continuous functions is always also a continuous function irrespective of whether we consider continuity at a given point or in a given interval. Does this property of finite sums also hold for infinite series? If all terms of the series

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots \quad (1)$$

are continuous in the interval (a, b) and if the series (1) is convergent at every point of this interval, then are we justified in maintaining that the sum $s(x)$ of this series is also continuous in the interval (a, b) ? We already know that this is in general not so. In § 73 we have considered the following series as an example

$$u_1(x) = x, \quad u_n(x) = x^n - x^{n-1} \quad (n > 1),$$

all terms of which are continuous in the interval $(0, 1)$, and we found that this series is convergent at every point of the interval $(0, 1)$ and its sum is equal to

$$s(x) = \begin{cases} 0 & (0 \leq x < 1), \\ 1 & (x = 1), \end{cases}$$

i.e. it is a continuous function. However, let us note that we have constructed this example in order to obtain a non-uniformly convergent series and we have, in fact, found that this series is non uniformly convergent in the interval $(0, 1)$. We therefore naturally tend to think that this non-uniform convergence is responsible for discontinuity of the obtained sum and this phenomenon could not take place, had we constructed a uniformly convergent series. This assumption can be fully confirmed by the following theorem :

Theorem 1. *If all terms of the series (1) which is uniformly convergent in the interval (a, b) are continuous in that interval, then the sum $s(x)$ of the series (1) is continuous in the interval (a, b) .*

Owing to the fact that continuity of the terms $u_n(x)$ of the series (1) is completely equivalent to continuity of the partial sums $s_n(x)$ of this series, the statement of theorem 1 is equivalent to the statement that if all terms of the sequence $s_1(x), s_2(x), \dots, s_n(x), \dots$ which tend uniformly to the limiting function $s(x)$ in the interval (a, b) , are continuous in that interval, then the function $s(x)$ is also continuous in that interval.

Note. We have seen that the sequence of functions

$$f_1(x), f_2(x), \dots, f_n(x), \dots$$

converges uniformly to the function $f(x)$ in the interval (a, b) if for every $\varepsilon > 0$ a number n_0 can be found such that for $n \geq n_0$ and for $a \leq x \leq b$ we have :

$$|f_n(x) - f(x)| < \varepsilon.$$

It is evident that uniform convergence of the series (1) is equivalent to uniform convergence of the sequence $s_1(x), s_2(x), \dots, s_n(x), \dots$ of its partial sums.

Proof. Let ε be an arbitrary positive number and let α be an arbitrary point in the interval (a, b) . Since the series (1) is uniformly convergent in that interval, we have for sufficiently large values of n :

$$|s(x) - s_n(x)| < \frac{1}{3} \varepsilon \quad (a \leq x \leq b). \quad (2)$$

Let us now fix a definite number n which satisfies this inequality. Since the function $s_n(x)$ is continuous at the point α , there would exist a $\delta > 0$ such that

$$|s_n(x) - s_n(\alpha)| < \frac{1}{3} \varepsilon, \quad (3)$$

if and only if $|x - \alpha| < \delta$. But

$$|s(x) - s(\alpha)| = |[s(x) - s_n(x)] + [s_n(x) - s_n(\alpha)] + [s_n(\alpha) - s(\alpha)]| \leq \\ \leq |s(x) - s_n(x)| + |s_n(x) - s_n(\alpha)| + |s_n(\alpha) - s(\alpha)|;$$

it follows from (2) that the first and third terms on the right-hand side are less than $\varepsilon/3$ irrespective of the points x and α in (a, b) ; it follows from (3) that the second term is less than $\varepsilon/3$, provided $|x - \alpha| < \delta$. Hence, provided this condition holds, each of the three terms on the right-hand side is less than $\varepsilon/3$ and therefore their sum is less than ε ; we thus obtain:

$$|s(x) - s(\alpha)| < \varepsilon,$$

if $|x - \alpha| < \delta$; since $\varepsilon > 0$ is arbitrary, the function $s(x)$ is continuous at the point α ; and owing to the fact that α can be any point in (a, b) , the function $s(x)$ is continuous in that interval; theorem 1 is thus proved.

Uniform convergence of a series of continuous functions guarantees continuity of the sum of that series. In fact, in most concrete cases continuity of the sum is established by means of this method by making reference to continuity of the series itself. However, it is also interesting to note that in some cases a non-uniformly convergent series of continuous functions can have a continuous sum so that the converse of theorem 1 is not true. We shall now prove this by an example. Let us select the terms of the series (1) so that

$$s_n(x) = x^n(1 - x^n) \quad (n = 1, 2, \dots);$$

in this connection it is evidently sufficient to assume that

$$u_1(x) = s_1(x) = x(1 - x), \\ u_n(x) = s_n(x) - s_{n-1}(x) = x^n(1 - x^n) - x^{n-1}(1 - x^{n-1}) \quad (n > 1).$$

And since for $0 \leq x \leq 1$

$$0 \leq s_n(x) < x^n,$$

therefore $s_n(x) \rightarrow 0$ ($n \rightarrow \infty$), where $0 \leq x < 1$; and since $s_n(1) = 0$ for every n , therefore $s_n(1) \rightarrow 0$ ($n \rightarrow \infty$) so that $s_n(x)$ tends to zero at every point in the interval $(0, 1)$. The sum of the series (1) is identically zero and therefore continuous. But on the other hand we have for $x = 1/2$:

$$|r_n(x)| = s_n(x) = \frac{1}{4},$$

and therefore no matter what n is, the inequality

$$|r_n(x)| < \varepsilon$$

for $\varepsilon \leq 1/4$ cannot hold for all points x in the interval $(0, 1)$ so that the series is non-uniformly convergent. In this case it is also interesting to imagine this phenomenon visually. Fig. 48 represents the graphs of the functions $y = s_n(x)$ when $n = 1$, $n = 2$ and when n is very large; each of these functions has (as the reader can readily calculate) the maximum value $1/4$ which the function $s_n(x)$ takes when $x = 1/\sqrt[n]{2}$. Hence as n increases this maximum, whose value remains constant, will move to the right and tend to the point 1. Therefore, no matter how close to unity we select our point x , this maximum will move farther to the right than x for a sufficiently large n while at the point x the function $s_n(x)$ will decrease and tend to zero as n increases; on the other hand, however, no matter how large the value of n be, a point can be found ($x = 1/\sqrt[n]{2}$) where $s_n(x) = 1/4$, i.e. points can always be found where the tendency of $s_n(x)$ to zero will considerably "lag" behind and we are unable to find a value of n for which $s_n(x)$ would be less than, say $1/8$ at every point in the interval $(0, 1)$; this shows nonuniform convergence of the constructed series.

Hence if uniform convergence of a series of continuous functions is in general not a necessary condition for continuity of its sum, a very important class of functions exists, for continuity of whose sums this condition is necessary. They are series with non-negative terms (and, in general, series with constant signs). In fact, let us assume that the series (1) has a continuous sum $s(x)$ in the interval (a, b) ; let all $u_n(x)$ be continuous in (a, b) and

$$u_n(x) \geq 0 \quad (n = 1, 2, \dots; a \leq x \leq b).$$

Let ε be an arbitrary positive number; a suffix n can be found for every point x in the interval (a, b) such that $r_n(x) = s(x) - s_n(x) < \varepsilon$; and since the function $r_n(x)$ is evidently continuous, this equation which holds for the point x should, according to the lemma § 23, also remain valid in an interval which contains the point x within itself (or when this point is one of its ends, i.e. $x = a$ or $x = b$). The

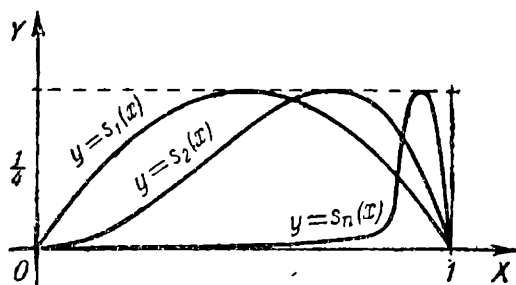


Fig. 48.

set of all such subintervals constructed for all the points x in the interval (a, b) will evidently cover this line. It follows from the theorem on finite coverage (lemma 2 § 18) that a finite sequence $\Delta_1, \Delta_2, \dots, \Delta_s$ of constructed subintervals exists which also covers the

given interval (a, b) . It follows from the construction of the covering system of subintervals that for every subinterval Δ_k ($1 \leq k \leq s$) there exists a suffix n_k such that

$$r_{n_k}(x) < \varepsilon$$

for all points x in the subinterval Δ_k . But since all $u_n(x)$ are non-negative, therefore $r_n(x)$ can only decrease as n increases; hence if we denote by m the greatest of the s numbers n_1, n_2, \dots, n_s , then the inequality

$$r_m(x) < \varepsilon$$

will hold for all the points in every subinterval Δ_k and hence also for all points in the interval (a, b) . And since ε is arbitrary, this proves uniform convergence of the series (1) in the interval (a, b) .

Theorem 2. *For the series (1) whose terms are continuous and non-negative in the interval (a, b) uniform convergence is the necessary and sufficient condition in order that the sum of the series should also be continuous in the interval (a, b) .*

§ 75. Term-by-term integration and differentiation of series

We know from integral calculus that the sum of a finite number of functions integrable in the interval (a, b) is also integrable in that interval and integral of the sum is equal to the sum of integrals of individual terms. Can this rule be extended to infinite series? If all terms of the series

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots \quad (1)$$

are integrable in the interval (a, b) and if the series (1) converges at every point in this interval, can we maintain that the sum $s(x)$ of this series is integrable in the interval (a, b) and that.

$$\int_a^b s(x) dx = \int_a^b u_1(x) dx + \int_a^b u_2(x) dx + \dots + \int_a^b u_n(x) dx + \dots? \quad (2)$$

If the equation (2) holds, we shall say that the series (1) can be *integrated term-by-term* in the interval (a, b) . If we denote the partial sum and the remainder of the series (1) respectively by $s_n(x)$ and $r_n(x)$, we evidently have

$$\lim_{n \rightarrow \infty} \int_a^b s_n(x) dx = \int_a^b s(x) dx, \quad \lim_{n \rightarrow \infty} \int_a^b r_n(x) dx = 0$$

and, conversely, either of these two relations leads to the relation (2).

It can be readily shown that term-by-term integration of series of functions is not always possible. In future we shall assume for the sake of simplicity that the terms of the series (1) are continuous in (a, b) , for many different complications can occur even with this restriction. It may at first happen that the sum $s(x)$ of the series (1) is not integrable in the interval (a, b) . Let us consider the following example (in which we shall define the functions $s_n(x)$, as we know that the terms $u_n(x)$ of the series (1) are defined directly and uniquely with reference to them). Let us assume that

$$s_n(x) = \begin{cases} n^2x & (0 \leq x \leq \frac{1}{n}), \\ \frac{1}{x} & (\frac{1}{n} \leq x \leq 1); \end{cases}$$

the graph of the function $y = s_n(x)$ is shown in Fig. 49. For every $x > 0$ we have $s_n(x) = 1/x$, provided $1/n \leq x$; therefore $\lim_{n \rightarrow \infty} s_n(x) = 1/x$ for every $x > 0$. If $x = 0$, then $s_n(0) = 0$ for any n and therefore $\lim_{n \rightarrow \infty} s_n(0) = 0$. Hence the function $s_n(x)$ has for any x ($0 \leq x \leq 1$) the following quantity as its limit :

$$s(x) = \begin{cases} \frac{1}{x} & (0 < x \leq 1), \\ 0 & (x = 0); \end{cases}$$

in other words, the series (1) converges at every point in the interval $(0, 1)$ and its sum is the function $s(x)$. The partial sums $s_n(x)$ and therefore also the terms $u_n(x)$ of the series are continuous and can therefore be integrated in the interval $(0, 1)$. However, the function $s_n(x)$ cannot be integrated in that interval. In fact, since the function $s(x)$ is non-negative, we have for any α ($0 < \alpha < 1$)

$$\begin{aligned} \int_0^1 s(x) dx &\geq \int_\alpha^1 s(x) dx = \\ &= \int_\alpha^1 \frac{dx}{x} = \ln \frac{1}{\alpha} : \end{aligned}$$

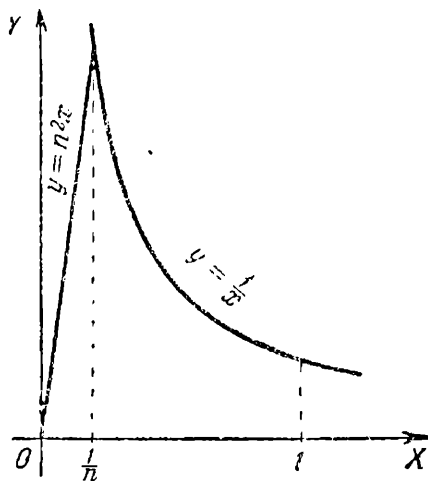


Fig. 49.

but $\ln 1/\alpha$ is as large as we please for a sufficiently small α and we arrive at an obvious contradiction.

It can happen on the other hand that the function $s(x)$ is integrable, but the series on the right-hand side of the equation (2) diverges. Let us select the functions $s_n(x)$ for $n \geq 2$ as follows: let $s_n(x) = 0$ for $x > 2/n$ and let $s_n(x)$ vary in the interval $(0, 2/n)$ as shown in Fig. 50. We have $s_n(0) = 0$ for any n ; if, however, $0 < x \leq 1$, we have $s_n(x) = 0$ for $2/n \leq x$ and therefore $\lim_{n \rightarrow \infty} s_n(x) = s(x) = 0$ at every point x in the interval $(0, 1)$ so that its sum is identically zero; but on the other hand, the integral

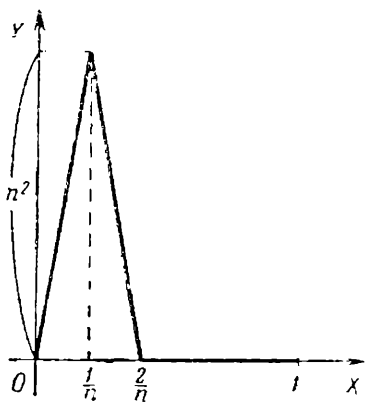


Fig. 50.

$$\int_0^1 s_n(x) dx$$

is equal to the area of an isosceles triangle shown in Fig. 50 and is therefore equal to n ; we therefore have (assuming for the sake of generality that $s_1(x) = u_1(x) = 0$ for $0 \leq x \leq 1$)

$$\int_0^1 s_n(x) dx = \sum_{k=1}^n \int_0^1 u_k(x) dx = n;$$

hence the sum increases indefinitely as $n \rightarrow \infty$ and the series on the right-hand side of the relation (2) diverges.

Finally it may happen that the sum $s(x)$ is integrable and the series on the right-hand side of (2) is convergent, but the equation (2) is not valid. To obtain an example of this kind it is sufficient to assume that the altitude of the triangle shown in Fig. 50 is equal to n instead of n^2 and retain the former definition of the function $s_n(x)$. We then have, as before, $s(x) = 0$ ($0 \leq x \leq 1$) and consequently

$$\int_0^1 s(x) dx = 0;$$

but now

$$\int_0^1 s_n(x) dx = \sum_{k=1}^n \int_0^1 u_k(x) dx = 1 \quad (n = 2, 3, \dots);$$

the right-hand side of the equation (2) is equal to unity and the left-hand side equal to zero.

We shall now show that neither of the cases considered above hold for uniformly convergent series which can therefore always be integrated term-by-term.

Theorem 1. *If all terms of the series (1) are continuous in the interval (a, b) and if this series converges uniformly in that interval, then the relation (2) holds.*

Proof. Uniform convergence of the series (1) at first implies continuity and therefore also integrability of its sum $s(x)$. Moreover, no matter how small $\varepsilon > 0$ be, we have for a sufficiently large n

$$|r_n(x)| < \varepsilon \quad (a \leq x \leq b).$$

Therefore for sufficiently large n

$$\left| \int_a^b r_n(x) dx \right| \leq \int_a^b |r_n(x)| dx \leq \varepsilon (b-a);$$

hence

$$\int_a^b r_n(x) dx \rightarrow 0 \quad (n \rightarrow \infty).$$

and this, as we have said before, is equivalent to the relation (2). Theorem 1 is thus proved.

Uniform convergence of a series of continuous functions which is a sufficient condition for term-by-term integration of the series is, however, not the necessary condition for this purpose. This can be readily shown with reference to the example considered in § 74 of a non-uniformly convergent series

$$s_n(x) = x^n - x^{2n}, \quad s(x) = 0 \quad (0 \leq x \leq 1).$$

Here,

$$\int_0^1 s_n(x) dx = \frac{1}{n+1} - \frac{1}{2n+1}, \quad \int_0^1 s(x) dx = 0,$$

and therefore

$$\lim_{n \rightarrow \infty} \int_0^1 s_n(x) dx = \int_0^1 s(x) dx;$$

this relation, as we know, is equivalent to the relation (2); hence the series can be integrated term-by-term, although its convergence is non-uniform.

Finally we shall make the following remark. If the terms of the series (1) are continuous and the series is uniformly convergent in the interval (a, b) , then the same condition also holds for any interval (a, x) , where $a < x < b$. We therefore have

$$\int_a^x s(y) dy = \sum_{n=1}^{\infty} \int_a^x u_n(y) dy. \quad (3)$$

The terms of the series on the right-hand side of this equation are functions of x , which are continuous in the interval (a, b) . Denoting the remainder of this series by $R_n(x)$ we evidently have:

$$R_n(x) = \int_a^x [s(y) - s_n(y)] dy = \int_a^x r_n(y) dy;$$

let n be so large that $|r_n(y)| < \varepsilon$ ($a \leq y \leq b$); in that case for $a \leq x \leq b$

$$|R_n(x)| \leq \int_a^x |r_n(y)| dy \leq \varepsilon (x - a) < \varepsilon (b - a).$$

This shows that the series (3) converges *uniformly* in the interval (a, b) . We thus arrive at the following proposition which can be regarded as a generalisation of theorem 1.

Theorem 2. *If all terms the series (1) are continuous in the interval (a, b) and the series converges uniformly in that interval, then the relation (3) holds uniformly for $a \leq x \leq b$.*

Finally let us consider the problem of term-by-term differentiation of series of functions. The implications of this problem are already clear to us. The sum of a finite number of functions differentiable at a given point x is also differentiable at that point and

the derivative of the sum is equal to the sum of derivatives of individual terms. We want to know, under what circumstances this condition can be extended to infinite series of functions.

Let the series (1) converge at every point in the interval (a, b) and let all terms of this series have continuous derivatives in that interval. Let us assume that the series

$$u_1'(x) + u_2'(x) + \dots + u_n'(x) + \dots \quad (4)$$

is uniformly convergent in the interval (a, b) . Let us denote by $s(x)$ the sum of the series (1) and by $t(x)$ the sum of the series (4). It follows from theorem 2 that we have for $a \leq x \leq b$:

$$\begin{aligned} \int_a^x t(y) dy &= \sum_{n=1}^{\infty} \int_a^x u_n'(y) dy = \sum_{n=1}^{\infty} [u_n(x) - u_n(a)] = \\ &= \sum_{n=1}^{\infty} u_n(x) - \sum_{n=1}^{\infty} u_n(a) = s(x) - s(a). \end{aligned}$$

It follows from a well-known property of integrals that the left-hand side of this equation is differentiable with respect to x and its derivative is equal to $t(x)$; we therefore conclude that the function $s(x)$ is differentiable and

$$s'(x) = t(x) = \sum_{n=1}^{\infty} u_n'(x) \quad (a \leq x \leq b),$$

i.e. the series (1) can be differentiated term-by-term at the point x . We have thus proved the following proposition:

Theorem 3. *Let the series (1) converge at every point in the interval (a, b) and let its sum be equal to $s(x)$ ($a \leq x \leq b$). If all terms of this series have continuous derivatives in the interval (a, b) and if the series (4) is uniformly convergent in that interval, then the function $s(x)$ also has a continuous derivative in the interval (a, b) and*

$$s'(x) = \sum_{n=1}^{\infty} u_n'(x) \quad (a \leq x \leq b),$$

i.e. the series (4) can be differentiated term-by-term at every point in the interval (a, b) .

Hence uniform convergence of the series (1) and not of the series (4) which is composed of derivatives of terms of the series (1) is, in this case, the necessary condition for term-by-term differentiation.

As a rule the condition established by theorem 3 is very convenient when applied to concrete series; in practice term-by-term differentiation of many series is established by means of this condition. It must be noted that differentiability of the sum $s(x)$ of the given series is not assumed in theorem 3 but proved on the basis of the conditions of the theorem.

For exercises to § 75 *cf.* Problem Book by B. P. Demidovich, Section V, Nos. 287, 294, 295.

CHAPTER X.

POWER SERIES AND SERIES OF POLYNOMIALS

§ 76. Region of convergence of a power series

Among various functions studied in mathematical analysis the simplest and theoretically and practically most important class is the so-called *power series*, i.e. series of the type

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n + \dots, \quad (1)$$

where a_0, a_1, \dots, a_n are constant real numbers; sometimes a more general expression of the following kind is also called a power series :

$$a_0 + a_1(x - a) + a_2(x - a)^2 + \dots + a_n(x - a)^n + \dots,$$

where a is a constant real number. The importance and simplicity of this type of series is first of all due to the fact that the partial sums $s_n(x)$ of the power series are expressed in terms of ordinary polynomials; therefore if the series (1) is convergent, its sum $s(x)$ which is, in general, a very complicated function, can be expressed approximately by a polynomial; the accuracy of this approximation can be as high as we please, provided we take a polynomial of a sufficiently high degree (i.e. the partial sum $s_n(x)$ with a sufficiently large n).

As with any other series of functions, we must at first study the *region of convergence* of this series, i.e. the values of x for which this series will converge and the values for which it will diverge. We have noticed in the previous paragraph that the region of convergence of some series of functions can be represented by a set of very complicated structure; we shall now learn that region of convergence of a *power series* always has a very simple form which considerably simplifies the study of this class of series.

The general form of region of convergence of a power series depends on the following important property of series of this class.

Theorem 1. *If the series (1) converges for $x = \alpha$, then it converges absolutely for every value of x for which*

$$|x| < |\alpha|.$$

Proof. It follows from our assumption on convergence of the series

$$a_0 + a_1\alpha + a_2\alpha^2 + \dots + a_n\alpha^n + \dots$$

that $a_n\alpha^n \rightarrow 0$ for $n \rightarrow \infty$; this, in its turn, implies existence of a number c so that

$$|a_n\alpha^n| < c \quad (n = 1, 2, \dots).$$

Let us now assume that $|x| < |\alpha|$. In that case when $n \geq 1$

$$|a_n x^n| = |a_n \alpha^n| \cdot \left| \frac{x}{\alpha} \right|^n < c \left| \frac{x}{\alpha} \right|^n.$$

Since the progression

$$\sum_{n=1}^{\infty} \left| \frac{x}{\alpha} \right|^n$$

is convergent, it follows from theorem 1 § 68 that the series

$$\sum_{n=1}^{\infty} |a_n x^n|,$$

is also convergent and this implies absolute convergence of the series (1). The theorem is thus proved.

The geometrical picture which illustrates theorem 1 implies that if the power series converges at a point α on the number line, then it will converge absolutely at any point closer to 0 than α . Let us now find the form of region of convergence of a power series in the light of theorem 1.

1. Every power series (1) converges at $x = 0$, for at that point all terms except the first are equal to zero. Hence the point $x = 0$ belongs to the region of convergence of every power series. Can it

happen that the series (1) diverges for any $x \neq 0$, i.e. its region of convergence consists of a single point $x = 0$? The series

$$1 + x + 2^2 x^2 + 3^3 x^3 + \dots + n^n x^n + \dots$$

shows that this is possible; in fact, if $x \neq 0$, then for $n > 1/|x|$

$$|n^n x^n| = |nx|^n > 1;$$

the n -th term of the series $n^n x^n$ does not tend to zero for $n \rightarrow \infty$ and the series diverges. Hence the region of convergence of a power series can consist of a single point $x = 0$.

2. The opposite limiting case arises when the series (1) converges for every value of x , i.e. when the whole number line is its region of convergence; the fact that this case is also possible can be seen from the series

$$1 + x + \frac{x^2}{2^2} + \frac{x^3}{3^3} + \dots + \frac{x^n}{n^n} + \dots;$$

since we have for $n > 2|x|$

$$|x|^n < \frac{n^n}{2^n}, \quad \frac{|x|^n}{n^n} < \frac{1}{2^n},$$

therefore for every value of x the terms of this series have a smaller absolute value than the corresponding terms of a convergent geometrical progression for all sufficiently large n ; this, as a result of the principle of comparison of series, proves convergence of the given series.

3. In all other cases there exist values of $x \neq 0$ for which the series (1) converges and other values for which it diverges. Let us at first prove that the region of convergence of the series (1) is, in this case, a bounded set. In fact, let α be an arbitrary point of divergence of the series (1); it then follows from theorem 1 that the series (1) should diverge for every value of x for which $|x| > |\alpha|$ and therefore each point of convergence of this series must satisfy the inequality

$$|x| \leq |\alpha|,$$

which proves that the points of the set of convergence are bounded.

Let us denote by r the upper bound of the set of points of convergence of the series (1), which exist as we have just proved. We say that the series (1) is absolutely convergent when $|x| < r$ and that

it diverges when $|x| > r$. The latter is evident from the definition of the number r . To prove the former let us assume that $|x| < r$ and that

$$r - |x| = \lambda > 0.$$

Since r is the upper bound of the set of points of convergence of the series (1), a point of convergence y can be found such that

$$y > r - \lambda = |x|.$$

It then follows from theorem 1 that the point x is a point of absolute convergence of the series (1), which was to be proved.

Hence if the region of convergence of the series (1) is not restricted to a single point $x = 0$ and if it does not include the whole number line, then an $r > 0$ always exists such that the series (1) converges for $|x| < r$ (i.e. in the interval $(-r, +r)$) and diverges for $|x| > r$ (i.e. outside that interval).

So as not to exclude the other two cases which we have considered above, it is always very convenient to consider in the first case the point 0 to correspond to the interval $(-r, +r)$ for $r = 0$ and in the second case the number line (or similar interval) for $r = +\infty$. Having accepted this we can formulate the result without exceptions.

Theorem 2. *There exists a number r ($0 \leq r \leq +\infty$) for every power series such that the series converges absolutely for $|x| < r$ and diverges for $|x| > r$.*

The number r is said to be *radius of convergence* and the interval $(-r, +r)$ *interval of convergence* of the given series; we shall consider somewhat later, whether this interval should be open or closed.

We can thus see that region of convergence of power series is always an interval with centre at the point 0; in some cases this interval may reduce to the single point $x = 0$ or it may include the whole number line. Hence one of the basic problems in the theory of power series is to determine radius of convergence of the series (1) from its "coefficients" a_n ($n = 1, 2, \dots$).

Modern science can solve this problem in the most general form, but we cannot consider this general solution here. We shall only consider one particular case which gives the desired result in many practical problems.

Theorem 3. *Let the coefficients of the series (1) be such that*

$$\left| \frac{a_{n+1}}{a_n} \right| \rightarrow l \text{ as } n \rightarrow \infty;$$

in that case

$$r = \begin{cases} \frac{1}{l}, & \text{if } 0 < l < +\infty, \\ \infty, & \text{if } l = 0, \\ 0, & \text{if } l = \infty. \end{cases}$$

Proof. Let $0 < l < +\infty$; in that case, as $n \rightarrow \infty$

$$\left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| = \left| \frac{a_{n+1}}{a_n} \right| \cdot |x| \rightarrow l |x|, \quad (2)$$

and

$$|x| < \frac{1}{l}, \quad l |x| < 1;$$

the series (1) converges absolutely at the point x in accordance with the test 2 (corollary) § 68. Conversely, if

$$|x| > \frac{1}{l}, \quad l |x| > 1,$$

the series $\sum_{n=1}^{\infty} |a_n x^n|$ diverges at the point x in accordance with the same test; it therefore follows that

$$r = \frac{1}{l}.$$

When $l = 0$, the relation (2) shows that for $n \rightarrow \infty$ and for every x

$$\left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| \rightarrow 0,$$

and therefore, in accordance with the same test, the series diverges for every x and $r = +\infty$.

Finally when $l = +\infty$, we have for every $x \neq 0$

$$\left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| \rightarrow \infty \quad (n \rightarrow \infty);$$

the series (1) diverges for every $x \neq 0$ and $r = 0$.

Example 1. Let us consider the series

$$\sum_{n=1}^{\infty} n^s x^n,$$

where s is an arbitrary real number. Since for $n \rightarrow \infty$

$$\frac{(n+1)^s}{n^s} = \left(1 + \frac{1}{n}\right)^s \rightarrow 1,$$

therefore it follows from theorem 3 that the radius of convergence r of our series is equal to unity for every s .

Example 2. Consider the series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

When $n \geq 1$,

$$\frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \frac{1}{n+1} \rightarrow 0 \quad (n \rightarrow \infty).$$

Hence the radius of convergence of the series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ is equal to

$r = +\infty$, i.e. the series converges for every value of x .

Should the interval of convergence of the given series be open or closed? In other words, does the given series converge or diverge at the points $x = -r$ and $x = r$?

These simple examples show that there is no single answer to these questions. Some series are convergent at both ends of the interval of convergence so that the *closed* interval $(-r, +r)$ serves as the region of convergence; other series, however, diverge when $x = r$ and $x = -r$ and therefore the *open* interval $(-r, +r)$ is the region of convergence for these series; finally there also exist other series which converge at one of the two ends of the interval of convergence and diverge at the other so that their region of convergence is the “*semi-open*” interval $(-r, +r)$.

Let us now consider corresponding examples.

Example 3. The series

$$1 + \frac{x}{1^2} + \frac{x^2}{2^2} + \dots + \frac{x^n}{n^2} + \dots$$

has, according to example 1, a radius of convergence equal to unity, and it is therefore absolutely convergent at both ends of the interval of convergence.

Example 4. The series

$$1 - \frac{x}{1} + \frac{x^2}{2} + \dots + \frac{x^n}{n} + \dots$$

has, according to example 1, a radius of convergence equal to unity. When $x = -1$, we obtain the Leibnitz series

$$1 - \frac{1}{1} + \frac{1}{2} - \frac{1}{3} + \dots,$$

which, as we know, is conditionally convergent. When $x = 1$, we obtain the (divergent) harmonic series. Hence the region of convergence of this series is the semi-open interval $(-1 \leq x < 1)$.

Example 5. The geometrical progression

$$1 + x + x^2 + \dots + x^n + \dots$$

has a radius of convergence equal to unity and diverges at both ends of the interval of convergence.

For further exercises cf. Problem Book by B.P. Demidovich, Section V, Nos. 125, 126, 132.

§ 77. Uniform convergence and its consequences

We have seen in the last chapter that *uniform* convergence has a great influence on different properties of series of functions. Having established the general form of the region of convergence of power series, we shall now naturally study uniformity of this convergence.

Can we say that any power series

$$a_0 + a_1x + \dots + a_nx^n + \dots \tag{1}$$

is uniformly convergent in the *open* interval $(-r, +r)$, where r is the radius of convergence of this series? We have already seen by an

example of a geometrical progression that this statement would not be generally valid. In fact, the series

$$1 + x + x^2 + \dots + x^n + \dots \quad (2)$$

has the open interval $(-1, +1)$ as its interval of convergence; the remainder of the series

$$r_n(x) = \sum_{k=n+1}^{\infty} x^k = \frac{x^{n+1}}{1-x}$$

tends to zero as $n \rightarrow \infty$ irrespective of x ($-1 < x < +1$); however, no matter how large n be, $r_n(x) \rightarrow \infty$ for $x \rightarrow 1$ and therefore no matter how n be, $r_n(x)$ will be as large as we please, provided x is close to unity; hence convergence of the series (2) in the open interval $(-1, +1)$ is non-uniform.

However, the power series converges uniformly in any interval which, together with its ends, lies within the interval of convergence, as is shown by the following theorem :

Theorem 1. *If r is the radius of convergence of the given series and if $0 < r' < r$, then the series (1) is uniformly convergent in the interval $(-r', +r')$.*

Proof. Since $r' < r$, the series (1) is absolutely convergent at the point $x = r'$, i.e., the following series converges

$$\sum_{n=1}^{\infty} |a_n| r'^n;$$

but when $|x| \leq r'$, we have :

$$|a_n x^n| \leq |a_n| r'^n \quad (n = 1, 2, \dots),$$

and it follows from theorem 2 § 73 that the series (1) converges uniformly in the interval $|x| \leq r'$; theorem 1 is thus proved.

This theorem has many corollaries which are all important in the theory and applications of power series. At first it follows that

Theorem 2. *The sum of a power series is continuous at every interior point of the interval of convergence.*

In fact, every interior point of the interval of convergence can be confined within an interval which, together with its ends, lies

within the interval of convergence. It follows from theorem 1 that the series will converge uniformly within this interval and, according to theorem 1 § 74, its sum is continuous.

Moreover, owing to the fact that a uniformly convergent series of continuous functions can always be integrated term-by-term (theorem 1 § 75), it follows from theorem 1 that

Theorem 3. *At any interior point x of the interval of convergence of the series (1)*

$$\int_0^x s(u) du = \sum_{n=0}^{\infty} a_n \int_0^x u^n du = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1},$$

where $s(x)$ denotes the sum of the series (1).

As we know (theorem 2 § 75), this last series is uniformly convergent in any interval in which the series (1) is uniformly convergent; hence it is also uniformly convergent in any interval which completely belongs to the interval of convergence of the series (1).

Finally, the following proposition which establishes the possibility of term-by-term differentiation of a power series within its interval of convergence is of fundamental importance in the theory of power series and in all its applications.

Theorem 4. *The sum $s(x)$ of the power series (1) is differentiable at every point inside the interval of convergence $(-r, +r)$ of this series; the series*

$$\sum_{n=1}^{\infty} n a_n x^{n-1}, \quad (3)$$

obtained as a result of term-by-term differentiation of the series (1), has the same radius of convergence r and its sum is equal to $s'(x)$ ($|x| < r$).

Proof. Let the numbers ρ and ρ' satisfy the inequality $0 < \rho < \rho' < r$. It follows from example 1 § 76 ($s = 1$) that the series

$$\sum_{n=1}^{\infty} n \lambda^n$$

is convergent for $0 \leq \lambda < 1$ so that assuming

$$\lambda = \frac{\rho}{\rho'}$$

we have :

$$n \left(\frac{\rho}{\rho'} \right)^n \rightarrow 0 \quad (n \rightarrow \infty),$$

and therefore there exists a number n independent of $c > 0$ such that

$$n \left(\frac{\rho}{\rho'} \right)^n < c \quad (n = 1, 2, \dots).$$

Having established this result we now note that

$$n |a_n| \rho^{n-1} = \frac{1}{\rho} n \left(\frac{\rho}{\rho'} \right)^n |a_n| \rho'^n < \frac{c}{\rho} |a_n| \rho'^n;$$

since $\rho' < r$, the series

$$\sum_{n=1}^{\infty} |a_n| \rho'^n$$

is convergent; but in this case the last inequality shows that the series (3) is also (absolutely) convergent for $x = \rho$. And since ρ can be as close to r as we please, the radius of convergence R of the series (3) is not less than r . Hence theorem 1 shows that the series (3) is uniformly convergent in any interval $-\rho \leq x \leq \rho$ if $0 < \rho < r$. But in this case we can maintain on the basis of the general theorem 3 § 75 that the function $s(x)$ has a derivative equal to the sum of the series (3) for $-r < x < r$. To conclude the proof of theorem 4 it only remains to show that $R = r$. This follows from the fact that the series (1) obtained as a result of term-by-term integration of the series (3) from 0 to x should, according to theorem 3, converge for $-R < x < R$; therefore $r \geq R$; and since we have already found that $R \geq r$, therefore $R = r$, and theorem 4 is proved.

Many important and far-reaching conclusions can be drawn from this theorem. At first this theorem shows that the sum $s(x)$ of a power series within its interval of convergence is always not only continuous but also differentiable. And since in this case the function $s'(x)$ appears to be the sum of a power series with the same interval of convergence $(-r, +r)$, we can again apply theorem 4 to this function. This shows that the second derivative $s''(x)$ of the function $s(x)$ exists at every interior point of the interval $(-r, +r)$ and is the sum of a power series resulting from the second term-by-term differentiation of the series (1). It is obvious that this argument can be continued *ad infinitum* and it leads to the following general result:

Theorem 5. *If r is the radius of convergence of the series (1), then its sum $s(x)$ has derivatives of all orders at every interior point of the interval $(-r, +r)$ and the function $s^{(n)}(x)$ ($n = 1, 2, \dots$) is the sum of the power series obtained as a result of n term-by-term differentiations of the series (1) and has the same radius of convergence r :*

$$s^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1) \dots (k-n+1) a_k x^{k-n} \quad (-r < x < +r). \quad (4)$$

For exercises to § 77 cf. Problem Book by B.P. Demidovich, Section V, Nos. 191, 192, 226, 200.

§ 78. Expansion of functions into power series

We have so far only investigated series which were given to us; we have found their region of convergence and studied the properties of their sums. However, in applications we are frequently dealing with the converse problem; we are given a function $s(x)$ and it is required to find if this function can be sum of a power series in the given interval (or, as it is usually said, if it can be “expanded into a power series”); if an expansion is possible, the question arises to find the coefficients of this series and determine its radius of convergence. We shall deal with these problems in this paragraph.

At first theorem 5 § 77 shows that the function $s(x)$ can be expanded into a power series if this function has derivatives of all orders at every point in the given interval (which we can assume to be of the form $(-r, +r)$, $r > 0$). This considerably narrows down the class of functions which can be expanded into power series; however, we must keep in mind that all elementary functions satisfy this requirement and, therefore, from a practical point of view, this condition is not so restrictive. Let us assume that the function $s(x)$ satisfies this requirement, i.e. there exist $s^{(n)}(x)$ for every $n \geq 0$ ($s^{(0)}(x) = s(x)$) and for every x ($-r < x < +r$). If $s(x)$ can be expanded into a power series

$$s(x) = \sum_{k=0}^{\infty} a_k x^k, \quad (1)$$

then, as we know, the relation (4) § 77 would hold for the function $s^{(n)}(x)$ ($n = 0, 1, 2, \dots$). If $x = 0$, this relation gives:

$$s^{(n)}(0) = n! a_n \quad (n = 0, 1, 2, \dots);$$

hence

$$a_n = \frac{s^{(n)}(0)}{n!} \quad (n = 0, 1, 2, \dots). \quad (2)$$

Thus the coefficients a_n of the power series whose sum is the function $s(x)$ are uniquely determined in terms of this function by means of formula (2).

This result is of great theoretical and practical importance. It shows from the theoretical point of view that every function $s(x)$ can have only one power series whose sum in an interval is equal to $s(x)$; in other words, two power series which converge in a given interval always have different sums in that interval. From the practical point of view this result enables us to calculate readily and easily the coefficients of the power series represented by $s(x)$; evaluation of derivatives of all orders of this function at the point $x = 0$ is only necessary for this purpose.

Hence if the function $s(x)$ can be expanded into the power series (1), this series always has the form

$$s(x) = \sum_{k=0}^{\infty} \frac{s^{(k)}(0)}{k!} x^k. \quad (3)$$

The series (3) for this function is known as the *Maclaurin series* of this function regardless of its region of convergence and irrespective of whether its sum does or does not coincide with the function $s(x)$. Hence every function with derivatives of all orders at $x = 0$ has a Maclaurin series; evidently this does not yet solve the problem of expanding the function $s(x)$ into a power series, since 1) the series (3) may be divergent for every $x \neq 0$ and 2) if this series converges, its sum may be a function other than $s(x)$. All that we know so far is restricted by the fact that *if the function $s(x)$ can, in general, be expanded into a power series, this series should be its Maclaurin series.*

However, this restricted result is very important. Until we have obtained this result, we could not find expansion of the function $s(x)$ into a power series, for we had no information about the coefficients of a possible series of this kind; now we are already studying a definite concrete series (3) and we must find its region of convergence and also find whether its sum coincides with the function $s(x)$ in this region.

We have already met the partial sums $s_n(x)$ of the Maclaurin series in Chapter IX and at that time, as also now, we were interested in finding the difference $s(x) - s_n(x)$. However, our new problem differs essentially from the old one. In Chapter IX we were not dealing with infinite series; we were only interested in finding the difference $s(x) - s_n(x)$ for a constant n and a sufficiently small x , and for this purpose we developed different special expressions for the quantity $s(x) - s_n(x) = r_n(x)$ which we called the "last term" of the Maclaurin formula. Now we are mainly interested in convergence of the series (3) with regard to the function $s(x)$, *i.e.* we are interested in finding the same last term $r_n(x)$ for the given x and $n \rightarrow \infty$. The special expressions for the last term developed in Chapter IX can also be used in this new problem. We shall have many examples of this kind later. At present we must emphasize again that in order to establish possibility of expanding the function $s(x)$ into a power series we cannot be satisfied by existence of derivatives of all orders for this function alone but we must also prove that the Maclaurin series is convergent (which is very easy); we are compelled to study the behaviour of the difference

$$r_n(x) = s(x) - \left[s(0) + \frac{s'(0)}{1!}x + \frac{s''(0)}{2!}x^2 + \dots + \frac{s^{(n)}(0)}{n!}x^n \right]$$

as $n \rightarrow \infty$. In fact, it can happen that the Maclaurin series constructed for the function $s(x)$ is convergent, but its sum is other than $s(x)$. For this purpose let us consider again the function which we have considered in § 41:

$$\varphi(x) = \begin{cases} e^{-1/x^2} & (x \neq 0), \\ 0 & (x = 0), \end{cases}$$

for which $\varphi^{(n)}(0) = 0$ ($n = 0, 1, 2, \dots$). If the function $s(x)$ can be expanded into the power series (1) (which, as we know, coincides with its Maclaurin series), then the function

$$s^*(x) = s(x) + \alpha \varphi(x),$$

where α is an arbitrary constant real number, evidently has the same Maclaurin series as the function $s(x)$; according to our proposition the sum of this series is equal to $s(x)$ and is therefore other than $s^*(x)$ (if $\alpha \neq 0$). We have seen above that the power series into which our function can be expanded is uniquely determined (as its Maclaurin series); now we see that, conversely, *one and the same series can serve as the Maclaurin series for an infinite number of different functions*. If the sum of the series is equal to one of the functions of this family,

then the Maclaurin series of any other function $f(x)$ belonging to the same family will be convergent, but its sum will be other than $f(x)$.

Let us finally remark that we have mentioned at the beginning of this chapter series of the following type :

$$a_0 + a_1(x-a) + a_2(x-a)^2 + \dots + a_n(x-a)^n + \dots, \quad (4)$$

where a is a constant real number; the substitution of the variable $x = a + y$ transforms this series into the simpler form of a power series

$$\sum_{n=0}^{\infty} a_n y^n;$$

therefore all properties of power series established above can be extended with slight modifications to series of the type (4). The region of convergence of the series (4) is always an (open, closed or semi-open) interval of the form $(a-r, a+r)$ ($0 \leq r \leq +\infty$). If $s(x)$ is the sum of the series (4), then $s^{(n)}(x)$ exists for every $n \geq 0$ and for every x ($a-r < x < a+r$) and

$$a_n = \frac{s^{(n)}(a)}{n!} \quad (n = 0, 1, 2, \dots),$$

so that the series (4) is *Taylor's series* for the function $s(x)$:

$$s(x) = s(a) + \frac{s'(a)}{1!}(x-a) + \frac{s''(a)}{2!}(x-a)^2 + \dots + \frac{s^{(n)}(a)}{n!}(x-a)^n + \dots$$

We shall now consider how some other important elementary functions can be expanded into power series. In many cases the possibility of this expansion is established by means of the following general proposition :

Theorem 1. *If there exists a positive number such that*

$$|s^{(n)}(x)| < C \quad (-r \leq x \leq r, n = 0, 1, 2, \dots),$$

then the function $s(x)$ can be expanded into a power series in the interval $-r \leq x \leq r$.

Proof. We have seen § 39 that the last term of the Maclaurin series can be represented in the form

$$r_n(x) = \frac{x^{n+1}}{(n+1)!} s^{(n+1)}(\theta x) \quad (0 < \theta < 1).$$

Hence when $-r \leq x \leq r$,

$$|r_n(x)| \leq \frac{Cr^{n+1}}{(n+1)!}.$$

But for any $r > 0$ we have $\frac{r^n}{n!} \rightarrow 0$ ($n \rightarrow \infty$); this follows directly, say, from the convergence of the series

$$\sum_{n=0}^{\infty} \frac{r^n}{n!}.$$

(example 5 § 76). Therefore when $-r \leq x \leq r$,

$$r_n(x) \rightarrow 0 \quad (n \rightarrow \infty),$$

and, consequently,

$$s(x) = \sum_{n=0}^{\infty} \frac{s^{(n)}(0)}{n!} x^n \quad (-r \leq x \leq r),$$

which proves theorem 1.

For the functions $s(x) = \sin x$ and $s(x) = \cos x$ we have :

$$|s^{(n)}(x)| \leq 1 \quad (-\infty < x < +\infty, \quad n = 0, 1, 2, \dots);$$

these functions can therefore be expanded into power series which converge along the whole number line.

For the functions $s(x) = e^x$ we have in any interval $-r \leq x \leq r$

$$|s^{(n)}(x)| = e^x \leq e^r \quad (n = 0, 1, 2, \dots);$$

hence the functions e^x can be expanded into a power series for $-r \leq x \leq r$ and therefore also along the whole number line, since the number $r > 0$ is arbitrary.

We know from § 39 the coefficients of the Maclaurin series for the functions $\sin x$, $\cos x$ and e^x ; therefore we can simply write :

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \dots,$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots + (-1)^n \frac{x^{2n}}{(2n)!} + \dots,$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots,$$

$$(-\infty < x < +\infty).$$

For the function $s(x) = 1/(1+x)$ we have for $x > -1$

$$|s^{(n)}(x)| = \frac{n!}{(1+x)^{n+1}} \rightarrow \infty \quad (n \rightarrow \infty),$$

and theorem 1 does not hold. We know, however, that the Maclaurin series for the function $s(x)$

$$1 - x + x^2 - x^3 + \dots$$

has a radius of convergence equal to unity and its sum is equal to $s(x)$ for $|x| < 1$. Owing to the fact that for $x > -1$

$$\int_0^x \frac{du}{1+u} = \ln(1+x),$$

it follows from theorem 3 § 77 that we have for $-1 < x < 1$

$$\begin{aligned} \ln(1+x) &= \int_0^x \frac{du}{1+u} = \sum_{n=0}^{\infty} (-1)^n \int_0^x u^n du = \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n+1} = \\ &= \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{n-1} \frac{x^n}{n} + \dots; \end{aligned}$$

the radius of convergence of this series is equal to unity (cf. example 4 § 76); hence the function $\ln(1+x)$ can be expanded into a power series only in the interval $(-1, +1)$. Similarly the series

$$\frac{1}{1+x^2} = \sum_{n=0}^{\infty} (-1)^n x^{2n},$$

whose radius of convergence is unity, can by means of integration be expanded into the following series

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1} = \frac{x}{1} - \frac{x^3}{3} + \frac{x^5}{5} - \dots + (-1)^n \frac{x^{2n+1}}{2n+1} + \dots$$

which is also convergent in the interval $(-1, +1)$.

We have obtained expansions of the functions $\ln(1+x)$ and $\arctan x$ into power series by means of the same method and both these series converge only in the interval $(-1, +1)$. However, there is an essential difference between them. The fact that the expansion

of the function $\ln(1+x)$ cannot be continued beyond the interval $(-1, +1)$ is intelligible, since the function $\ln(1+x) \rightarrow -\infty$ for $x \rightarrow -1$ and $\ln(1+x)$ is void for $x \leq -1$; on the other hand, the function $\arctan x$ is defined and has derivatives of all orders along the whole number line; nevertheless, its expansion into the Maclaurin series is possible only in the interval $(-1, +1)$.

Finally, let us consider expansion of the function $s(x) = (1+x)^\alpha$ into a power series, where α is an arbitrary constant real number. We have :

$$s^{(n)}(0) = \alpha(\alpha-1) \dots (\alpha-n+1),$$

which gives us the following expression for the coefficients of the Maclaurin series for the function $s(x)$:

$$a_n = \frac{\alpha(\alpha-1) \dots (\alpha-n+1)}{n!} \quad (n = 0, 1, 2, \dots).$$

If α is zero or a natural number, then all a_n vanish from a certain number onwards, and the resulting Maclaurin series simply coincides with Newton's binomial formula. However, for all other values of α the coefficients a_n are non-zero and we obtain an infinite series. We can evidently restrict ourselves and only consider this case.

Since

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{\alpha-n}{n+1} \right| \rightarrow 1 \quad (n \rightarrow \infty),$$

therefore, according to theorem 3 § 76, the radius of convergence of the Maclaurin series for the function $s(x)$ is equal to unity. Hence beyond the interval $(-1, +1)$ this function cannot be expanded into a power series. We will now show that expansion is possible in this interval, i.e. when $|x| < 1$,

$$1 + \sum_{n=1}^{\infty} \frac{\alpha(\alpha-1) \dots (\alpha-n+1)}{n!} x^n = (1+x)^\alpha.$$

For this purpose we shall use the expression for the last term of the Maclaurin series obtained in § 39 :

$$r_n(x) = \frac{(1-\theta)^n x^{n+1}}{n!} = s^{(n+1)}(\theta x),$$

where $0 < \theta < 1$. In our case

$$s^{(n+1)}(x) = \alpha(\alpha-1) \dots (\alpha-n)(1+x)^{\alpha-n-1},$$

and consequently

$$\begin{aligned} r_n(x) &= \frac{(1-\theta)^n x^{n+1}}{n!} \alpha(\alpha-1) \dots (\alpha-n)(1+\theta x)^{\alpha-n-1} = \\ &= \frac{(\alpha-1)(\alpha-2) \dots (\alpha-1-n+1)}{n!} x^n \cdot \alpha x (1+\theta x)^{\alpha-1} \left(\frac{1-\theta}{1+\theta x} \right)^n. \end{aligned}$$

Since $x > -1$, therefore $0 < 1-\theta < 1+\theta x$ and

$$0 < \frac{1-\theta}{1+\theta x} < 1;$$

further, θ only depends on n in the expression $\alpha x (1+\theta x)^{\alpha-1}$; but since we always have $0 < \theta < 1$, the expression $|\alpha x (1+\theta x)^{\alpha-1}|$ is always confined between the positive numbers

$$\alpha x |1+|x||^{\alpha-1} \text{ and } |\alpha x| (1-|x|)^{\alpha-1},$$

which are independent of n ; hence denoting by k the larger of these two numbers we have for every n

$$|\alpha x (1+\theta x)^{\alpha-1}| \leq k.$$

We thus obtain the evaluation

$$|r_n(x)| \leq k \left| \frac{(\alpha-1)(\alpha-2) \dots (\alpha-1-n+1)}{n!} x^n \right|;$$

on the right-hand side the coefficient of k represents the absolute value of the n -th term of the Maclaurin series for the function $(1+x)^{\alpha-1}$; but we have shown that this series converges for every index if $|x| < 1$; therefore the n -th term of this series should tend to zero as $n \rightarrow \infty$ and we obtain

$$r_n(x) \rightarrow 0 \quad (n \rightarrow \infty),$$

which was to be proved.

In conclusion to this paragraph we shall review once again (in a condensed form) the expansion of several important transcendental functions into Maclaurin series :

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (-\infty < x < +\infty).$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \quad (-\infty < x < +\infty).$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \quad (-\infty < x < +\infty).$$

$$\ln(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} \quad (-1 < x \leq 1)^*.$$

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \quad (-1 \leq x \leq 1)^*.$$

$$(1+x)^\alpha = 1 + \sum_{n=1}^{\infty} \frac{\alpha(\alpha-1) \dots (\alpha-n+1)}{n!} x^n \quad (-1 < x < 1)^*.$$

These expansions occur frequently in applications and they must be remembered by heart like tables of derivatives or tables of primitives of simple functions.

For exercises to § 78 cf. Problem Book by B. P. Demidovich, Section V, Nos. 140-143, 149-154, 163, 169, 171, 173, 178, 187, 189.

§ 79. Series of polynomials

We have already said that one of the main advantages of expanding functions into power series is the approximate representation of these functions in terms of polynomials. In fact, if an arbitrary function $f(x)$ can be expanded into a power series which converges uniformly to $f(x)$ in the interval (a, b) , then for $\varepsilon > 0$, which can be as small as we please, and for a sufficiently large n we have

$$|f(x) - s_n(x)| < \varepsilon \quad (a \leq x \leq b),$$

where $s_n(x)$ are the partial sums of the power series, which are therefore polynomials. In this connection it is important to note that the theory of power series not only enables us to establish possibility of the approximate replacement of functions by polynomials but also makes it possible to find these polynomials by expressing their coefficients in terms of the function $f(x)$ and its derivatives at $x = 0$.

*) The behaviour of the series shown above at the ends of the interval of convergence can be established by a more detailed investigation which we do not give here.

Thus if the expansion of a function into a power series makes it possible to express it approximately in terms of polynomials of a sufficiently high degree with as great an accuracy as we please, it is equally interesting to find if the converse proposition is also true; we already know that a comparatively restricted class of functions can be expanded into power series; for example, such functions must have derivatives of all orders, and even this restrictive condition is not sufficient. If uniform approximation in terms of polynomials with as high a degree of accuracy as we please could only be applied to functions which can be expanded into power series, the scope of these approximations would become very restricted.

We shall now agree to say that *the function $f(x)$ permits uniform approximation in terms of polynomials in the interval (a, b)* if there exists a polynomial $P(x)$ for arbitrarily small $\epsilon > 0$ such that

$$|f(x) - P(x)| < \epsilon \quad (a \leq x \leq b).$$

We have already seen that a function which can be expanded into a power series permits uniform approximation in terms of polynomials in every interval which lies entirely within the interval of convergence of the given series. However, a power series is a particular case of a series of a more general type

$$\sum_{n=1}^{\infty} P_n(x), \tag{1}$$

whose terms are composed of arbitrary polynomials $P_n(x)$. Let us assume that the series (1) converges uniformly in the interval (a, b) ; let us denote its sum by $f(x)$ and its partial sums by $s_n(x)$. According to the definition of uniform convergence we have a value of n for every $\epsilon > 0$ such that

$$|f(x) - s_n(x)| < \epsilon \quad (a \leq x \leq b),$$

and since $s_n(x)$ is the sum of a finite number of polynomials, it is also a polynomial, and therefore it implies that the function $f(x)$ permits uniform approximation in terms of polynomials in the interval (a, b) . Hence a function which can be expanded in a subinterval into a uniformly convergent series of polynomials permits uniform approximation in terms of polynomials in that interval. However, it can be readily seen that the converse proposition is also true. In fact, let the function $f(x)$ permit uniform approximation in terms of

polynomials in the interval (a, b) . In that case for every natural number n a polynomial $Q_n(x)$ can be found such that

$$|f(x) - Q_n(x)| < \frac{1}{n} \quad (a \leq x \leq b). \quad (2)$$

Let us assume that

$$P_1(x) = Q_1(x), P_n(x) = Q_n(x) - Q_{n-1}(x) \quad (n > 1);$$

in that case the partial sums of the series

$$\sum_{n=1}^{\infty} P_n(x)$$

will be expressed by the polynomials $Q_n(x)$ and, the inequality (2) shows that this series converges uniformly in the interval (a, b) and its sum is equal to $f(x)$.

Hence uniform approximation of the function $f(x)$ in terms of polynomials in the interval (a, b) is equivalent to the fact that the function $f(x)$ can be expanded into a uniformly convergent series of polynomials in that interval. We must now establish which functions permit this expansion.

At first it is obvious that this function must be *continuous* in the interval (a, b) ; in fact, since all polynomials are continuous functions, therefore (according to theorem 1 § 74) the sum of a uniformly convergent series of polynomials in the interval (a, b) must also be continuous in that interval. One of the most fundamental theorems of mathematical analysis is the converse fact discovered in the second half of the last century (by the German mathematician Weierstrass) that *every function which is continuous in the interval (a, b) permits uniform approximation in terms of polynomials in that interval (or, what is the same, an expansion into a uniformly convergent series of polynomials)*. Hence by passing from power series to the more general type of series of polynomials we considerably enlarge the class of functions which can be expanded into such series; if it was earlier necessary that a function should have derivatives of all orders, it is now no longer necessary to assume existence of even the first derivative.

The theoretical and practical applications of Weierstrass' theorem are so great that many different proofs were proposed for it from the time of its discovery; these proofs can be divided into

two groups: one group uses the properties of one or other special analytical apparatus in order to establish the theorem, while proofs of the second group are based on general considerations. Both groups are very instructive, for the arguments used in this connection find many applications in other analytical problems. In the next paragraph we give one of the simplest proofs of the first group proposed by Academician S. N. Bernstein.

§ 80. Theorem of Weierstrass

Theorem. *The function $f(x)$ which is continuous in the interval (a, b) permits uniform approximation in terms of polynomials in that interval.*

Proof. 1. All proofs of this basic theorem which rest on a special analytical apparatus require the preliminary elucidation of the properties of this apparatus. Therefore we must also in this case begin by giving the proof of an auxiliary elementary algebraic inequality.

Lemma. *For every natural number n and for every x the following inequality holds*)*

$$\sum_{k=0}^n C_n^k (k - nx)^2 x^k (1 - x)^{n-k} \leq \frac{n}{4}.$$

Proof. Newton's binomial formula gives us identically with respect to z :

$$\sum_{k=0}^n C_n^k z^k = (1 + z)^n; \quad (1)$$

differentiating with respect to z and multiplying by z this gives

$$\sum_{k=0}^n k C_n^k z^k = nz (1 + z)^{n-1}; \quad (2)$$

the repetition of the same operation gives:

$$\sum_{k=0}^n k^2 C_n^k z^k = nz (1 + z)^{n-1} + n(n-1)z^2 (1 + z)^{n-2} =$$

*) Here and in future we shall assume that $x^k = 1$ for $x = 0$ and $k = 0$; similarly when $x = 1$ and $k = n$, we assume that $(1 - x)^{n-k} = 1$. The numbers C_n^k are binomial coefficients with their usual combination notation.

$$\begin{aligned}
&= nz (1 + z)^{n-2} \{1 + z + (n - 1)z\} = \\
&= nz (1 + nz)(1 + z)^{n-2};
\end{aligned} \tag{3}$$

assuming in the relations (1), (2) and (3) that

$$z = \frac{x}{1 - x}$$

and multiplying by $(1 - x)^n$ we obtain for $x \neq 1$:

$$\sum_{k=0}^n C_n^k x^k (1 - x)^{n-k} = 1, \tag{4}$$

$$\sum_{k=0}^n k C_n^k x^k (1 - x)^{n-k} = nx, \tag{5}$$

$$\sum_{k=0}^n k^2 C_n^k x^k (1 - x)^{n-k} = nx (1 - x + nx), \tag{6}$$

and a test shows that the equations (4), (5) and (6) are also valid for $x = 1$.

Let us now add term by term the identities (4), (5) and (6), having multiplied the first by $n^2 x^2$ and the second by $-2nx$. This gives :

$$\begin{aligned}
\sum_{k=0}^n (n^2 x^2 - 2n x k + k^2) C_n^k x^k (1 - x)^{n-k} &= \\
&= n^2 x^2 - 2n^2 x^2 + nx (1 - x + nx),
\end{aligned}$$

or

$$\sum_{k=0}^n (k - nx)^2 C_n^k x^k (1 - x)^{n-k} = nx(1 - x),$$

and since for every real x

$$x(1 - x) \leq \frac{1}{4},$$

therefore the lemma is proved.

2. We shall now proceed to prove the theorem and begin by assuming that the given interval (a, b) is the interval $(0, 1)$; we assume that for every natural number n we have :

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k} = B_n(x);$$

$B_n(x)$ is evidently a polynomial of a degree not higher than n^*). We shall now note that $B_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ uniformly in the interval $(0, 1)$.

Let us denote by M the upper bound of the function $|f(x)|$ in the interval $(0, 1)$. Let ε be an arbitrary positive number so that the function $f(x)$ is continuous (and therefore also uniformly continuous) in the interval $(0, 1)$; there exists a positive number δ in this case such that when

$$|x_1 - x_2| \leq \delta, \quad 0 \leq x_1 \leq 1, \quad 0 \leq x_2 \leq 1,$$

we have

$$|f(x_1) - f(x_2)| < \varepsilon.$$

Our immediate problem is to find the difference $|B_n(x) - f(x)|$ for $0 \leq x \leq 1$. It follows from formula (4) that for every x

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k} = B_n(x),$$

which makes it possible to write the difference $B_n(x) - f(x)$ in a form convenient for evaluation

$$|B_n(x) - f(x)| = \sum_{k=0}^n \left[f\left(\frac{k}{n}\right) - f(x) \right] C_n^k x^k (1-x)^{n-k};$$

hence for $0 \leq x \leq 1$

$$|B_n(x) - f(x)| \leq \sum_{k=0}^n \left| f\left(\frac{k}{n}\right) - f(x) \right| C_n^k x^k (1-x)^{n-k}. \quad (7)$$

*These "Bernstein polynomials" $B_n(x)$ make the special analytical apparatus whose properties are used in the above proof.

Let us divide all the numbers k ($0 \leq k \leq n$) into two groups : the group (A) contains the numbers k for which

$$\left| \frac{k}{n} - x \right| \leq \delta, \quad (\text{A})$$

and the group (B) contains those numbers k for which

$$\left| \frac{k}{n} - x \right| > \delta; \quad (\text{B})$$

hence in the inequality (7) $\sum_{k=0}^n$ is correspondingly broken up into two

sums which we denote respectively by Σ_A and Σ_B . According to (A) we have in every term of Σ_A

$$\left| f\left(\frac{k}{n}\right) - f(x) \right| < \varepsilon,$$

and therefore

$$\Sigma_A \leq \varepsilon \sum_{k \in A} C_n^k x^k (1-x)^{n-k} \leq \varepsilon \sum_{k=0}^n C_n^k x^k (1-x)^{n-k} = \varepsilon. \quad (8)$$

But in every term of the sum Σ_B we have :

$$(k - nx)^2 > n^2 \delta^2, \quad \left| f\left(\frac{k}{n}\right) - f(x) \right| \leq 2M,$$

and therefore

$$\begin{aligned} \Sigma_B &\leq \frac{2M}{n^2 \delta^2} \sum_{k \in B} (k - nx)^2 C_n^k x^k (1-x)^{n-k} \leq \\ &\leq \frac{2M}{n^2 \delta^2} \sum_{k=0}^n (k - nx)^2 C_n^k x^k (1-x)^{n-k}; \end{aligned}$$

hence it follows from the above lemma that

$$\Sigma_B < \frac{M}{2n\delta^2}. \quad (9)$$

Finally we obtain from (7), (8) and (9) :

$$|B_n(x) - f(x)| \leq \Sigma_A + \Sigma_B < \varepsilon + \frac{M}{2n\delta^2};$$

if n is so large that $\frac{M}{2n\delta^2} < \epsilon$, then

$$|B_n(x) - f(x)| < 2\epsilon \quad (0 \leq x \leq 1).$$

and since $\epsilon > 0$ is arbitrarily small, we have uniformly in the interval $(0, 1)$

$$B_n(x) \rightarrow f(x) \quad (n \rightarrow \infty),$$

which was to be proved.

3. The extension of this theorem proved for the interval $(0, 1)$ to an arbitrary interval (a, b) ($a < b$) involves no further difficulties. Let the function $f(x)$ be continuous in the interval (a, b) . If we assume that $x = a + (b - a)y$, then $y = (x - a)/(b - a)$ so that y traverses the interval $(0, 1)$ while x traverses the interval (a, b) . Let us assume that

$$f(x) = f[a + (b - a)y] = \varphi(y) \quad (0 \leq y \leq 1).$$

It is evident that the function $\varphi(y)$ is continuous in the interval $(0, 1)$. Therefore no matter how small $\epsilon > 0$ be, it follows from the proved theorem that there exists a polynomial $P(y)$ such that

$$|\varphi(y) - P(y)| < \epsilon \quad (0 \leq y \leq 1),$$

or, what is the same,

$$\left| f(x) - P\left(\frac{x-a}{b-a}\right) \right| < \epsilon \quad (a \leq x \leq b);$$

but $P((x - a)/(b - a))$ is a polynomial of the variable x which for the sake of brevity can be denoted by $Q(x)$ so that

$$|f(x) - Q(x)| < \epsilon \quad (a \leq x \leq b);$$

since ϵ is an arbitrarily small number, this implies uniform approximation of the function $f(x)$ in terms of polynomials in the interval (a, b) . This proves the fundamental theorem completely.

We already know that this theorem is equivalent to the following statement: *every function $f(x)$ which is continuous in the interval (a, b) can be expanded into a series of polynomials which is uniformly convergent in that interval.*

CHAPTER XXI

TRIGONOMETRICAL SERIES

§ 81. Fourier coefficients

In this chapter we shall consider the theory of the so-called *trigonometrical series*; this is a class of series, which, after power series, is most important both in theory and application. A *trigonometrical series* is a series of the type

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx), \quad (1)$$

where $a_0, a_1, a_2, \dots, b_1, b_2, \dots$ are constant real numbers called the *coefficients* of the series (1). We shall see in this chapter that the properties of trigonometrical series are very different from those of power series and they are in some respects a more complicated and bulky apparatus for the study of functions they represent; however, in other respects they possess many advantages over power series. We cannot generally say which of these two classes of functions deserves precedence. The answer depends at first on the form of the function which is being studied and also on the problems which have to be solved with regard to this function. We already know that in order to represent a function by a power series it is necessary that this function should have derivatives of all orders; on the other hand, in order to expand a trigonometrical function into a power series it is sufficient, as we shall soon learn, that the first derivative should exist and be continuous*); thus the class of functions consisting of trigonometrical series is much wider than that of power series and this considerably increases the significance of trigonometrical series; on the

*) and even this condition is not always necessary.

other hand, as we know, the region of convergence of a power series always has a very simple form; it is an interval with centre at the point O . But the region of convergence of a trigonometrical series is generally a set of a very complicated structure and very sensitive methods are needed for its study, which we cannot consider in this book. In this respect power series have an advantage over trigonometrical series.

If we increase or decrease the variable x by 2π , then all terms of the series (1) evidently remain unchanged; if the series is convergent, its sum will not be affected, hence the sum of the series (1) is always a periodic function with a period 2π ; if the function $f(x)$ does not possess this periodicity, it can be represented by a trigonometrical series only in an interval less than 2π . This restriction is not necessary and can be readily avoided by a very simple method, as we shall see later. On the other hand, while studying trigonometrical series periodicity of its terms evidently enables us to confine ourselves to an arbitrary interval 2π in length; in such cases we usually take the interval $-\pi \leq x \leq \pi$.

The system of functions

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \dots \quad (2)$$

which forms the basis of every series of the type (1), possesses one remarkable property which holds the main key to the theory of trigonometrical series and is the source of almost all advantages of this apparatus. This property is due to the fact that any two functions of the system (2) are *mutually orthogonal* in every interval 2π in length. The functions $f_1(x)$ and $f_2(x)$ are said to be *mutually orthogonal in the interval (a, b)* if

$$\int_a^b f_1(x) f_2(x) dx = 0.$$

If we have a (finite or infinite) system of functions in which any two functions are mutually orthogonal in the interval (a, b) , then this system is said to be an *orthogonal system* in that interval. In order to prove that the system of functions (2) is orthogonal in every interval 2π in the length, it is evidently sufficient to show that

$$\int_{-\pi}^{\pi} \cos mx \cos nx dx = 0 \quad (m \neq n; m, n = 0, 1, 2, \dots),$$

$$\int_{-\pi}^{\pi} \sin mx \sin nx \, dx = 0 \quad (m \neq n; m, n = 1, 2, \dots),$$

$$\int_{-\pi}^{\pi} \cos mx \sin nx \, dx = 0 \quad (m = 0, 1, 2, \dots; n = 1, 2, \dots);$$

but

$$\cos mx \cos nx = \frac{1}{2} \{ \cos (m+n)x + \cos (m-n)x \},$$

$$\sin mx \sin nx = \frac{1}{2} \{ \cos (m-n)x - \cos (m+n)x \},$$

$$\cos mx \sin nx = \frac{1}{2} \{ \sin (n+m)x + \sin (n-m)x \}.$$

Hence the above three integrals can be written in the form of the following integrals

$$\int_{-\pi}^{\pi} \cos kx \, dx, \quad \int_{-\pi}^{\pi} \sin lx \, dx,$$

where the integer k is non-zero and the integer l is arbitrary; both integrals vanish, as can be readily shown by a simple calculation.

The property of orthogonality of systems of functions makes them a very convenient instrument in mathematical analysis, so that functions much more complicated than those entering the system (2) can be conveniently studied provided they form an orthogonal system. Modern science knows and uses many orthogonal systems and their theory is usually constructed on the lines of the theory of the system (2) and related trigonometrical series.

In order to demonstrate the first simple application of orthogonality of the system (2) we shall now consider a problem for trigonometrical series which is analogous to a problem solved earlier for power series: *assuming that the function $f(x)$ can be expanded into the trigonometrical series (1), find the coefficients a_k, b_k of this series.*

Let us assume that the series (1) is *uniformly* convergent in the interval $(-\pi, \pi)$ and owing to periodicity also on the whole number line and that its sum is equal to $f(x)$ so that we have

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \quad (-\pi \leq x \leq \pi). \quad (3)$$

Let n be an arbitrary natural number. Let us multiply all terms of the series (3) by the same function $\cos nx$; the new series so obtained also converges uniformly in the interval $(-\pi, \pi)$, since the remainder of this series is equal to $r_k(x) \cos nx$ and its absolute value does not exceed the remainder $r_k(x)$ of the series (3); the sum of the new series is evidently equal to $f(x) \cos nx$, so that we have :

$$f(x) \cos nx = \frac{a_0}{2} \cos nx + \sum_{k=1}^{\infty} (a_k \cos kx \cos nx + b_k \sin kx \cos nx) \\ (-\pi \leq x \leq \pi).$$

It follows from theorem 1 § 75 that this series can be integrated term-by-term in the interval $(-\pi, \pi)$. The integral of the left hand side is

$$\int_{-\pi}^{\pi} f(x) \cos nx \, dx.$$

As a result of orthogonality all the integrals on the right-hand side are equal to zero except the integral

$$\int_{-\pi}^{\pi} a_n \cos^2 nx \, dx$$

of the term of the series for which $k = n$; hence as a result of term-by-term integration we obtain

$$\int_{-\pi}^{\pi} f(x) \cos nx \, dx = \int_{-\pi}^{\pi} a_n \cos^2 nx \, dx. \quad (4)$$

But $\cos^2 nx = 1/2 (1 + \cos 2nx)$, therefore

$$\int_{-\pi}^{\pi} a_n \cos^2 nx \, dx = a_n \left\{ \frac{1}{2} \int_{-\pi}^{\pi} dx + \frac{1}{2} \int_{-\pi}^{\pi} \cos 2nx \, dx \right\} = \pi a_n.$$

Thus the equation (4) gives :

$$\int_{-\pi}^{\pi} f(x) \cos nx \, dx = \pi a_n,$$

and consequently

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx \quad (n = 1, 2, \dots). \quad (5)$$

Similarly by multiplying all terms of the series (3) by $\sin nx$ and integrating term-by-term in the interval $(-\pi, \pi)$ we obtain

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx \quad (n = 1, 2, \dots). \quad (6)$$

Finally, term-by-term integration of the series (3) itself in that same interval gives :

$$\int_{-\pi}^{\pi} f(x) \, dx = \pi a_0,$$

and therefore

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \, dx; \quad (7)$$

formula (7) can be regarded as a particular case of formula (5) for $n = 0$; in order to emphasize this generality the constant term of a trigonometrical series is usually denoted by $a_0/2$ and not by a_0 .

The formulae (5), (6), (7) completely solve our problem, for they enable us to find the coefficients a_k and b_k when the given function $f(x)$ (the sum of the trigonometrical series) is known, provided the series is uniformly convergent.

As with power series, we can see that these coefficients are *uniquely* expressed in terms of the function $f(x)$; hence there is only one uniformly convergent trigonometrical series whose sum is equal to the given function $f(x)$. In contrast to power series where the expressions of coefficients require existence of derivatives of all orders of the function $f(x)$, we can now see that the formulae (5), (6) and (7) are not required to be expanded any more beyond its properties mentioned

in the statement of our problem; in fact, it follows from uniform convergence of the series (3) (which we have assumed from the beginning) that the function $f(x)$ is continuous (and hence also integrable) (this also implies integrability of the functions $f(x) \cos nx$ and $f(x) \sin nx$); in order that formulae (5), (6) and (7) should be valid nothing but this integrability is required.

The numbers a_n, b_n expressed in terms of the function $f(x)$ by means of the formulae (5), (6), (7) are usually called *Fourier coefficients* of this function (in spite of the fact that the formulae (5) (6) and (7) were founded by Euler long before Fourier). Hence for every function $f(x)$ which is continuous in the interval $(-\pi, \pi)$ we can, by using these formulae, make the complete series of Fourier coefficients $a_0, a_1, a_2, \dots, b_1, b_2, \dots$; in this event the series (1) in which the numbers so determined serve as coefficients is called the *Fourier series* of the function $f(x)$; in short, every function $f(x)$ which is continuous in the interval $(-\pi, \pi)$ has a Fourier series. However, as in the case of power series, no conclusions can be drawn from the possibility of expanding the function $f(x)$ into a trigonometrical series. The Fourier series for the function $f(x)$ may be divergent for some (or even all) values of x . Moreover, even if we assume that this series is divergent for all values of x , we have no grounds to believe that its sum coincides with the function $f(x)$. Hence the question of what properties the function $f(x)$ must possess in order that it should be the sum of its Fourier series requires special study. At present we only know one thing: if a trigonometrical series does, in fact, exist and converge uniformly to the function $f(x)$ in the interval $(-\pi, \pi)$, then this series must be the Fourier series of this function.

In this course we cannot consider other orthogonal systems of functions than the system (2). It is nevertheless interesting to note that all we have said above with regard to coefficients and Fourier series is based purely on the property of orthogonality of the system (2) and is quite independent of any special characteristics of the trigonometrical functions which form the system; for this reason everything we have said can refer to every orthogonal system. If the continuous functions

$$\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x), \dots \quad (8)$$

form an orthogonal system in an interval (a, b) and if the series

$$\sum_{n=1}^{\infty} a_n \varphi_n(x) = f(x), \quad (9)$$

where a_n are constant real numbers, is uniformly convergent in the interval (a, b) , then, in the same way as above, we readily obtain :

$$\int_a^b f(x) \varphi_n(x) dx = a_n \int_a^b \varphi_n^2(x) dx \quad (n = 1, 2, \dots);$$

assuming, for the sake of simplicity, that*)

$$\int_a^b \varphi_n^2(x) dx = 1 \quad (n = 1, 2, \dots),$$

we therefore obtain :

$$a_n = \int_a^b f(x) \varphi_n(x) dx \quad (n = 1, 2, \dots). \quad (10)$$

The numbers a_n which we have found for the function $f(x)$ by means of the formula (10) are called the Fourier coefficients of this function and the series (9) its Fourier series (with respect to the orthogonal system (8)).

The general theory of orthogonal systems is one of the most important chapters in mathematical analysis and has numerous practical applications. At present many investigations are being made in this field. Much progress in this direction was made by our scientists Chebyshev and Liapunov and several other Soviet mathematicians (Bari, Kolgomorov, Luzin, Menshov, Steklov and others).

For exercises to § 81 cf. Problem Book by B. P. Demidovich, Section V, Nos. 332, 334, 341-344.

§ 82. Average approximation

Before proceeding to the solution of the fundamental problem of convergence of Fourier series we will now show that the Fourier coefficients of the given function possess great practical values regardless of the fact whether the series (3) § 81 is divergent or convergent. If this series converges at a point x , then the function

*) The system (8) which satisfies this condition is called *normalised*; evidently every system (8) can be normalised by multiplying its constituent functions by some constant numbers; thus in the system (2) it is sufficient to multiply the first term by $1/\sqrt{2\pi}$ and the remaining terms by $1/\sqrt{\pi}$.

$f(x)$ can be represented approximately with any required degree of accuracy by its partial sum:

$$s_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx),$$

which is a so-called "trigonometrical polynomial". In general, a trigonometrical polynomial of degree n is a sum of the form

$$T_n(x) = \frac{\alpha_0}{2} + \sum_{k=1}^n (\alpha_k \cos kx + \beta_k \sin kx), \quad (1)$$

where α_k, β_k are constant real numbers.

But we have seen in chapter 20 that regardless of the possibility of expanding a function into a power series there may arise the question of its approximate expression in terms of a polynomial. Similarly in this case, regardless of the possibility of expanding the function $f(x)$ into a trigonometrical series, we can also consider its approximate expression by a trigonometrical polynomial of the type (1).

If the order n of this trigonometrical polynomial is preassigned, then naturally the question arises how to select the coefficients α_k, β_k of the polynomial $T_n(x)$ so as to obtain the best approximation. If we are in this case considering the approximation of the function $f(x)$ not at one particular point but in the whole interval $(-\pi, \pi)$, then we must also define what is meant by a "best approximation". The difference $f(x) - T_n(x)$ whose magnitude (*i.e.* absolute value) we naturally regard as a measure of the quality of the given approximation, will have different values at different points of the interval $(-\pi, \pi)$. If we have two different trigonometrical polynomials $T_n(x)$, this difference will in general be smaller for the first of these polynomials at some points, while at other points it will be smaller for the second polynomial; we cannot directly see which of these two polynomials represents the function $f(x)$ better. To obtain a unique evaluation for the comparative quality of approximations given by different polynomials we must evidently agree to assess this quality in each case by a definite number. This number can be chosen by various methods in the same way as we can choose different thermometers for measuring temperatures; in this case, as in the above case, the advantage of this or other method of evaluation

does not depend so much on fundamental considerations but chiefly on convenience.

In the same way as we have evaluated the closeness of two points by the distance between them, so in this case in order to evaluate the closeness between the function $f(x)$ and the trigonometrical polynomial $T_n(x)$ we must clearly determine the "distance" between them; the smaller this distance is, the closer the approximation of the function $f(x)$ given by the polynomial $T_n(x)$. Naturally the determination of this distance must in one way or other take into consideration the magnitude of the difference $f(x) - T_n(x)$ at every point in the interval $(-\pi, \pi)$. One convenient definition of this distance is given by the upper bound of the quantity $|f(x) - T_n(x)|$ in the interval $(-\pi, \pi)$. Another possible definition can be provided by the "mean value" of the same quantity

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - T_n(x)| dx \quad (2)$$

in the interval $(-\pi, \pi)$. However, it is more convenient (and it is, in fact, most frequently used in the theory of orthogonal systems) to use the definition of the distance between $f(x)$ and $T_n(x)$ given by the mean value

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [f(x) - T_n(x)]^2 dx = \rho(f, T_n) \quad (3)$$

of the *square* of the difference $f(x) - T_n(x)$; this definition has a purely practical advantage over the definition (2): in calculations it is more convenient to deal with squares of functions than with their absolute values.

In future we shall define the distance between the function $f(x)$ and the trigonometrical polynomial by means of formula (3); we shall say that out of the two polynomials $T_{n,1}(x)$ and $T_{n,2}(x)$ the former gives a better approximation for the function $f(x)$ than the latter if

$$\rho(f, T_{n,1}) < \rho(f, T_{n,2});$$

the best approximation of all trigonometrical polynomials of order n is given by the polynomial $T_n(x)$, for which the distance $\rho(f, T_n)$ is the least.

Having thus agreed on this definition we are faced with a completely defined problem we must find the polynomial (1) of degree n , for which the quantity $\rho(f, T_n)$ has the least possible value. But finding of the polynomial $T_n(x)$ implies finding of its coefficients. The quantity $\rho(f, T_n)$ is evidently a function of these coefficients *i.e.* a function of $2n + 1$ numbers $\alpha_0, \alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ if the polynomial $T_n(x)$ is given by formula (1). Hence we must find those $2n + 1$ numbers for which the quantity $\rho(f, T_n)$ has its least possible value.

For this purpose let us represent the expression (3) for the quantity $\rho(f, T_n)$ in the form :

$$\frac{1}{2\pi} \left\{ \int_{-\pi}^{\pi} f^2(x) dx + \int_{-\pi}^{\pi} T_n^2(x) dx - 2 \int_{-\pi}^{\pi} f(x) T_n(x) dx \right\}. \quad (4)$$

Determining $T_n(x)$ by means of formula (1) and denoting by a_n, b_n the Fourier coefficients of the function $f(x)$ we obtain as a result of the formulae (5), (6), (7) § 81 :

$$\int_{-\pi}^{\pi} f(x) T_n(x) dx = \pi \left\{ \frac{\alpha_0 a_0}{2} + \sum_{k=1}^n (\alpha_k a_k + \beta_k b_k) \right\}. \quad (5)$$

On the other hand, bearing in mind the formula

$$\int_{-\pi}^{\pi} \cos^2 kx dx = \int_{-\pi}^{\pi} \sin^2 kx dx = \pi \quad (k = 1, 2, \dots)$$

and orthogonality of the system (2) § 81 we obtain :

$$\begin{aligned} \int_{-\pi}^{\pi} T_n^2(x) dx &= \int_{-\pi}^{\pi} \left\{ \frac{\alpha_0^2}{4} + \sum_{k=1}^n \alpha_k^2 \cos^2 kx + \beta_k^2 \sin^2 kx \right\} dx = \\ &= \pi \left\{ \frac{\alpha_0^2}{2} + \sum_{k=1}^n (\alpha_k^2 + \beta_k^2) \right\}. \end{aligned} \quad (6)$$

Substituting the expressions (5) and (6) in the expression (4) for the quantity $\rho(f, T_n)$ we have :

$$\begin{aligned} \rho(f, T_n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(x) dx + \frac{1}{2} \left\{ \frac{(\alpha_0^2 - 2\alpha_0 a_0)}{2} + \right. \\ &\quad \left. + \sum_{k=1}^n [(\alpha_k^2 - 2\alpha_k a_k) + (\beta_k^2 - 2\beta_k b_k)] \right\}; \end{aligned}$$

noting that

$$\alpha_k^2 - 2\alpha_k a_k = (\alpha_k - a_k)^2 - a_k^2,$$

$$\beta_k^2 - 2\beta_k b_k = (\beta_k - b_k)^2 - b_k^2,$$

we consequently obtain :

$$\begin{aligned} \rho(f, T_n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(x) dx - \frac{1}{2} \left\{ \frac{a_0^2}{2} + \sum_{k=1}^n (a_k^2 + b_k^2) \right\} + \\ &+ \frac{1}{2} \left\{ \frac{(\alpha_0 - a_0)^2}{2} + \sum_{k=1}^n [(\alpha_k - a_k)^2 + (\beta_k - b_k)^2] \right\}. \end{aligned}$$

The first two terms on the right-hand side of this equation are independent of the coefficients α_k, β_k of the polynomial $T_n(x)$; therefore the numbers α_k, β_k must be so chosen that the third term should have the least possible value, *i.e.* the quantity

$$\frac{1}{2} \left\{ \frac{(\alpha_0 - a_0)^2}{2} + \sum_{k=1}^n [(\alpha_k - a_k)^2 + (\beta_k - b_k)^2] \right\};$$

this quantity evidently vanishes if we choose

$$\alpha_k = a_k \quad (k = 0, 1, \dots),$$

$$\beta_k = b_k \quad (k = 1, 2, \dots),$$

and becomes positive for every choice of the numbers α_k, β_k . This solves our problem. We see that in order to obtain the best approximation for every n we must choose the corresponding Fourier coefficients of the function $f(x)$ for the coefficients of the polynomial $T_n(x)$. In doing so we must naturally bear in mind that this deduction will only be valid if we are evaluating the distance between the function $f(x)$ and the polynomial $T_n(x)$ by means of the quantity (3). If we use a different method for determining the distance, we obtain different values for the numbers α_k, β_k .

Approximations in which the distance between two functions is evaluated in terms of the average value of the square of their

difference are usually called *average approximations*. The result obtained above can therefore be formulated as follows :

Theorem 1. *From all trigonometrical polynomials of degree n the best average approximation is given by the following polynomial, provided the function $f(x)$ is continuous :*

$$T_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx),$$

where a_k, b_k are Fourier coefficients of the function $f(x)$. Also

$$\rho(f, T_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(x) dx - \frac{1}{2} \left\{ \frac{a_0^2}{2} + \sum_{k=1}^n (a_k^2 + b_k^2) \right\}. \quad (7)$$

The equation (7) gives one interesting result. Since

$$\rho(f, T_n) \geq 0,$$

therefore for every n

$$\frac{a_0^2}{2} + \sum_{k=1}^n (a_k^2 + b_k^2) \leq \frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x) dx;$$

the right-hand side of this equation is independent of n ; therefore the partial sums of the series

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty} (a_k^2 + b_k^2)$$

remain bounded for $n \rightarrow \infty$; and since this series has constant signs, it must be convergent. Hence the squares of the Fourier coefficients of a continuous function always form a convergent series. It also follows that when $n \rightarrow \infty$, we always have for a continuous function :

$$a_n \rightarrow 0, \quad b_n \rightarrow 0.$$

§ 83. Dirichlet-Liapunov theorem on closed trigonometrical systems

We have learnt in the theory of power series that the same Maclaurin series can be used for an infinite number of different functions. A simpler question naturally arises, and is equally

important, with regard to Fourier series; can the same trigonometrical series be the Fourier series of several different functions? We must at first draw attention to the fact that in solving this problem it is advisable to consider continuous functions only, for otherwise the resulting solution would be trivial; in fact, if we are to consider discontinuous functions, then we shall change the value of the given function $f(x)$ at a particular point; it can be readily shown that the new function will have the same Fourier coefficients as the function $f(x)$, since the integrals in the formulae (5), (6), (7) § 81 do not change in magnitude as a result of this change in the function $f(x)$.

The above problem, as we shall now see, is closely related to an important property of the orthogonal system

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \dots \quad (1)$$

of trigonometrical functions. Let us assume that there exist two different functions $f_1(x)$ and $f_2(x)$ which are continuous in the interval $(-\pi, \pi)$ and have the same Fourier series, or, what is the same thing, the same series of Fourier coefficients. Let us now consider the function

$$f(x) = f_1(x) - f_2(x);$$

it follows from the formulae (5), (6), (7) § 81 that any Fourier coefficient of the function $f(x)$ is equal to the difference between the corresponding coefficients of the functions $f_1(x)$ and $f_2(x)$ and is therefore equal to zero. We thus have :

$$\left. \begin{aligned} \int_{-\pi}^{\pi} f(x) \cos kx \, dx &= 0, \\ \int_{-\pi}^{\pi} f(x) \sin kx \, dx &= 0 \end{aligned} \right\} \quad (k = 0, 1, 2, \dots).$$

But this implies that the function $f(x)$ is orthogonal to any one of the functions of the system (1) in the interval $(-\pi, \pi)$. Therefore, *if there exist two different functions which are continuous in the interval $(-\pi, \pi)$ and have the same Fourier series, then there exists another function which is also continuous in the interval $(-\pi, \pi)$ but does not become identically zero; this function is orthogonal to all functions of the system (1) in that interval.*

It can be readily seen that the converse proposition also holds. In fact, if the function $f(x)$ is continuous and does not become

identically zero and is also orthogonal to all functions of the system (1) in the interval $(-\pi, \pi)$ and if $\varphi(x)$ is an arbitrary function which is continuous in that interval, then the function $\varphi(x) + f(x)$ will also be continuous in the interval $(-\pi, \pi)$; it does not coincide with $\varphi(x)$ whereas all its Fourier coefficients coincide with the corresponding coefficients of the function $\varphi(x)$.

The question therefore arises, whether it is permissible to "add" to the orthogonal system (1) another continuous function which is not identically zero so that the extended system should remain orthogonal. For this extension of the system (1), as we have just seen, it is necessary and sufficient that the two different functions which have the same Fourier series should exist. An orthogonal system which can be extended in this way is called an *open* system; a *closed* system, on the other hand, does not permit such an extension. We must therefore decide whether the orthogonal system (1) is closed or open.

The most important property of the closed system follows from the remarkable investigations by Dirichlet on convergence of trigonometrical series; the same facts were, however, earlier established and proved independently somewhat later by the outstanding Russian scientist Academician A. M. Liapunov. Therefore we shall in future call this theorem the Dirichlet-Liapunov theorem.

Dirichlet-Liapunov theorem. *The orthogonal system (1) is closed.*

Proof. Let the function $f(x)$ be continuous in the interval $(-\pi, \pi)$ and orthogonal to all functions of the system (1) in that interval. We must prove that $f(x) = 0$ $(-\pi \leq x \leq \pi)$.

It evidently follows from orthogonality of the function $f(x)$ and any other function of the system (1) that it is also orthogonal to an arbitrary trigonometrical polynomial $T(x)$. We shall prove converse of this theorem; let us assume that the function $f(x)$ is not identically zero in the interval $(-\pi, \pi)$, and on the basis of these considerations we shall construct a trigonometrical polynomial to which the given function cannot be orthogonal in the interval $(-\pi, \pi)$.

Let, for example, $f(x) > 0$ for $x = \alpha$, $-\pi < \alpha < \pi$. In that case we have $f(x) > 0$ for a sufficiently small $\delta > 0$ and for all points x in the interval $(\alpha - \delta, \alpha + \delta)$; let $c > 0$ be the smallest value of the function $f(x)$ in that interval so that

$$f(x) \geq c > 0 \quad (\alpha - \delta \leq x \leq \alpha + \delta).$$

Let us now assume that

$$T_n(x) = \left\{ \frac{1 + \cos(x - \alpha)}{2} \right\}^n,$$

where n is an arbitrary natural number. Raising the power in accordance with the binomial formula we evidently obtain:

$$T_n(x) = \sum_{r=0}^n c_r [\cos(x - \alpha)]^r,$$

where c_r are constant real numbers. It is known that for every $r \geq 0$ the function $(\cos x)^r$ can be represented as a linear combination of functions

$$1, \cos x, \cos 2x, \dots, \cos rx$$

with constant coefficients.*) Applying this expansion to all terms of the above sum we obtain:

$$T_n(x) = \sum_{r=0}^n d_r \cos r(x - \alpha),$$

where d_r are constants. Finally, since for every r

$$\cos r(x - \alpha) = \cos r\alpha \cos rx + \sin r\alpha \sin rx,$$

we obtain the following expression for $T_n(x)$:

$$T_n(x) = \frac{\alpha_0}{2} + \sum_{k=1}^n (\alpha_k \cos kx + \beta_k \sin kx),$$

*) Proof. This statement is obvious for $r = 1$. If

$$(\cos x)^r = \sum_{s=0}^r \alpha_s \cos sx,$$

then

$$\begin{aligned} (\cos x)^{r+1} &= (\cos x)^r \cos x = \sum_{s=0}^r \alpha_s \cos sx \cos x = \\ &= \frac{1}{2} \sum_{s=0}^r \alpha_s [\cos(s+1)x + \cos(s-1)x] = \sum_{s=0}^{r+1} \beta_s \cos sx. \end{aligned}$$

where α_k, β_k are constants. This shows that for every natural n the function $T_n(x)$ is a trigonometrical polynomial. We will now show that provided n is sufficiently large, this polynomial cannot be orthogonal to the function $f(x)$ in the interval $(-\pi, \pi)$.

Let us imagine the main outlines of the course of the function $T_n(x)$ in the interval $(-\pi, \pi)$ when n is large; this will also enable us to perceive clearly the true meaning which lies at the basis of the following proof. The quantity

$$\frac{1 + \cos(x - \alpha)}{2},$$

whose n -th degree is the polynomial $T_n(x)$, is evidently always non-negative; it is equal to unity for $x = \alpha$ and less than unity for all other values of x in the interval $(-\pi, \pi)$. Therefore when n is large, $T_n(x)$, which is always non-negative, is equal to unity for $x = \alpha$ and negligibly small for all other values of x which are only a small distance away from α , so that the course of the function approximately follows the graph shown in Fig. 51. We must prove that the integral

$$\int_{-\pi}^{\pi} f(x) T_n(x) dx \quad (2)$$

cannot be equal to zero when n is sufficiently large. For this purpose we can break the integral (2) into two parts:

$$\int_{\alpha-\delta}^{\alpha+\delta} f(x) T_n(x) dx = I_1$$

and

$$\left[\int_{-\pi}^{\alpha-\delta} + \int_{\alpha+\delta}^{\pi} \right] f(x) T_n(x) dx = I_2.$$

Since the second of these integrals includes the region in which $T_n(x)$ is negligibly small, we have reason to believe that the absolute value of the integral is also negligibly small. On the other hand, we know about the first integral that $f(x) \geq c > 0$ in it and the function $T_n(x)$ attains its maximum value. We therefore have good reasons to expect that the

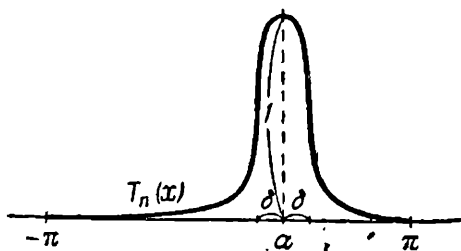


Fig. 51.

absolute value of the first integral will be considerably greater than the second integral; but this is all that is needed since if $I_1 > I_2$, we cannot have $I_1 + I_2 = 0$.

Let us now perform the necessary calculations. Owing to the fact that

$$\frac{1 + \cos(x - \alpha)}{2} = \cos^2\left(\frac{x - \alpha}{2}\right),$$

we have

$$I_1 = \int_{\alpha-\delta}^{\alpha+\delta} f(x) \cos^{2n}\left(\frac{x-\alpha}{2}\right) dx \geq c \int_{\alpha-\delta}^{\alpha+\delta} \cos^{2n}\left(\frac{x-\alpha}{2}\right) dx.$$

Assuming that $x = \alpha + y$ we obtain:

$$I_1 \geq c \int_{-\delta}^{\delta} \cos^{2n}\left(\frac{y}{2}\right) dy = 2c \int_0^{\delta} \left(1 - \sin^2 \frac{y}{2}\right)^n dy.$$

Since $0 \leq \sin y/2 < 1$ and $0 < \cos y/2 \leq 1$ in the interval $(0, \delta)$, therefore

$$\left(1 - \sin^2 \frac{y}{2}\right)^n \geq \left(1 - \sin \frac{y}{2}\right)^n \cos \frac{y}{2},$$

and we obtain:

$$I_1 \geq 2c \int_0^{\delta} \left(1 - \sin \frac{y}{2}\right)^n \cos \frac{y}{2} dy,$$

or, assuming that $\sin y/2 = z$,

$$I_1 \geq 4c \int_0^{\sin \frac{\delta}{2}} (1-z)^n dz = \frac{4c}{n+1} \left\{ 1 - \left(1 - \sin \frac{\delta}{2}\right)^{n+1} \right\} > \frac{2c}{n+1}, \quad (3)$$

so that the expression inside the braces is evidently greater than $1/2$ when n is sufficiently large.

We have evaluated the lower limit of the integral I_1 ; we shall now try to evaluate the upper limit of the absolute value of the integral I_2 . Let us denote by M the maximum value of the function

$|f(x)|$ in the interval $(-\pi, \pi)$. Within limits of the two integrals which form the integral I_2 (i.e. when $|x - \alpha| > \delta$), we have:

$$\cos^2\left(\frac{x - \alpha}{2}\right) \leq \cos^2\left(\frac{\delta}{2}\right)$$

and therefore

$$|T_n(x)| \leq \cos^{2n}\left(\frac{\delta}{2}\right);$$

hence

$$\begin{aligned} |I_2| &\leq M \cos^{2n}\left(\frac{\delta}{2}\right) [(\alpha - \delta) - (-\pi) + \pi - (\alpha + \delta)] < \\ &< 2\pi M \cos^{2n}\left(\frac{\delta}{2}\right) = 2\pi M r^n, \end{aligned} \quad (4)$$

where it is assumed that

$$r = \cos^2\left(\frac{\delta}{2}\right) < 1.$$

Owing to the fact that for a sufficiently large n we have *)

$$2\pi M r^n < \frac{2c}{n+1},$$

and it follows from (3) and (4) that $|I_2| < I_1$, provided n is sufficiently large. The equation

$$\int_{-\pi}^{\pi} f(x) T_n(x) dx = I_1 + I_2 = 0$$

cannot therefore hold and the Dirichlet-Liapunov theorem is proved.

We can thus see in the light of the problem considered above that the behaviour of Fourier series differs from that of Maclaurin series: every trigonometrical series can be Fourier series of only one continuous function.

§ 84. Convergence of Fourier series

We shall now consider the fundamental problem, viz. the properties which the function $f(x)$ must possess in order that its Fourier series should be convergent and its sum be equal to the given

*) This follows from the relation $nr^n \rightarrow 0$ ($n \rightarrow \infty$) (cf. proof of theorem 4 § 77).

function. This problem is, on the whole, very complicated and modern science has not yet succeeded in solving it fully. On the one hand, we know many tests which enable us to determine expansion of this or other class of functions into Fourier series; however, it has been shown by many examples that relatively simple functions have divergent Fourier series. In this paragraph we shall only prove one proposition which shows how wider is the class of functions which can be expanded into Fourier series than the class of functions which can be expanded into power series.

Theorem. *The function $f(x)$ with a period 2π , whose first derivative is continuous everywhere, can be expanded into a uniformly convergent trigonometrical series along the number line (it follows from the basic result of § 81 that this series is its Fourier series).*

Proof. Let us denote by a_k, b_k the Fourier coefficients of the function $f(x)$ and by a'_k, b'_k the Fourier coefficients of the function $f'(x)$. The formulae (5), (6) § 81 and integration by parts for $k > 0$ give :

$$\begin{aligned}\pi a_k &= \int_{-\pi}^{\pi} f(x) \cos kx \, dx = \\ &= \left(\frac{f(x) \sin kx}{k} \right) \Big|_{-\pi}^{\pi} - \frac{1}{k} \int_{-\pi}^{\pi} f'(x) \sin kx \, dx = -\frac{\pi b'_k}{k}, \\ \pi b_k &= \int_{-\pi}^{\pi} f(x) \sin kx \, dx = \\ &= \left(-\frac{f(x) \cos kx}{k} \right) \Big|_{-\pi}^{\pi} + \frac{1}{k} \int_{-\pi}^{\pi} f'(x) \cos kx \, dx = \frac{\pi a'_k}{k}.\end{aligned}$$

We therefore have :

$$a_k = -\frac{b'_k}{k}, \quad b_k = \frac{a'_k}{k} \quad (k = 1, 2, \dots). \quad (1)$$

Since for any two numbers α and β it follows from

$$\alpha^2 + \beta^2 - 2|\alpha\beta| = (|\alpha| - |\beta|)^2 \geq 0$$

that

$$2|\alpha\beta| \leq \alpha^2 + \beta^2,$$

therefore the relation (1) gives :

$$2|a_k| \leq b'_k{}^2 + \frac{1}{k^2}, \quad 2|b_k| \leq a'_k{}^2 + \frac{1}{k^2},$$

and hence for every x

$$|a_k \cos kx + b_k \sin kx| \leq |a_k| + |b_k| \leq \frac{1}{2}(a'_k{}^2 + b'_k{}^2) + \frac{1}{k^2}. \quad (2)$$

Since the numbers a'_k, b'_k are Fourier coefficients of the continuous function $f'(x)$, it follows from the final result of § 82 that the series

$$\sum_{k=1}^{\infty} (a'_k{}^2 + b'_k{}^2)$$

is convergent; and owing to the fact that the series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is also con-

vergent, the right-hand side of the inequality (2) represents the k -th term of a convergent numerical series with positive terms; but it also follows from (2) that the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \quad (3)$$

is absolutely and uniformly convergent along the whole number line. Let us denote its sum by $s(x)$; it follows from the fundamental result of § 81 that the series (3) is the Fourier series of the function $s(x)$ whereas, by definition, it is the Fourier series of the function $f(x)$. Since both functions $f(x)$ and $s(x)$ are continuous, therefore, it follows from § 83 that they must coincide; this proves our theorem.

§ 85. Generalised trigonometrical series

We have already mentioned in § 81 a self-evident fact that functions with a period 2π can only be expanded into trigonometrical series along the whole number line. A function $f(x)$ which does not possess this property, may, at best, be expanded only in (arbitrary) interval $(a, a + 2\pi)$ of length 2π ; it is then evidently also necessary that $f(a + 2\pi) = f(a)$. This restriction, if it could not readily be removed, would greatly reduce the use of trigonometrical series. In

this paragraph we shall briefly consider how it is possible to enlarge the concept of trigonometrical series simply and naturally by removing this obstacle. For the sake of simplicity we shall always assume that the given function $f(x)$ has a continuous derivative at every point in the given interval (a, b) in which we wish to expand it into a trigonometrical series.

If we are considering an interval $(a, a + \lambda)$ which is shorter than 2π ($\lambda < 2\pi$), then our problem is evidently very simple; all we have to do is to continue the function $f(x)$ in the interval $(a + \lambda, a + 2\pi)$ in an arbitrary manner and we have $f(a + 2\pi) = f(a)$ and $f'(a + 2\pi) = f'(a)$; the function $f(x)$ would then have a continuous derivative in the whole interval $(a, a + 2\pi)$ (this can, of course, be done in an infinite number of ways). It follows from § 84 that the continued function can be represented in the interval $(a, a + 2\pi)$ by a uniformly convergent series whose sum evidently coincides with the given function $f(x)$ in the interval $(a, a + \lambda)$; this solves our problem. It is also interesting to note in this connection that when $\lambda < 2\pi$, continuation of the function $f(x)$ along the whole interval $(a, a + 2\pi)$, which can be done in an infinite number of ways, will also give an infinite number of different Fourier series for the initial function. The sums of these Fourier series will obviously be different functions if we consider them in the whole interval $(a, a + \pi)$. However, they will all coincide with the function $f(x)$ in the interval $(a, a + \lambda)$. Thus the function can in general be represented by an infinite number of different trigonometrical series in an interval of length $< 2\pi$.

Let us now assume that $\lambda > 2\pi$; we again assume that the function $f(x)$ has a continuous derivative in the interval $(a, a + \lambda)$ and $f(a + \lambda) = f(a)$, $f'(a + \lambda) = f'(a)$; we can then assume that the function $f(x)$ is periodically continued across the ends of the interval $(a, a + \lambda)$ (with a period λ) along the whole number line.

Let us assume that

$$x = a + \frac{\lambda}{2\pi} y, \quad f(x) = f\left(a + \frac{\lambda}{2\pi} y\right) = \varphi(y).$$

The function $\varphi(y)$ is evidently a periodic function with a period 2π (since on increasing y by 2π we evidently increase x by λ) and it has a continuous derivative for every y ; denoting by a_k and b_k the Fourier coefficients of this function we obtain for every y , as a result of the theorem in § 84,

$$\varphi(y) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos ky + b_k \sin ky),$$

where the series is uniformly convergent along the whole number line. Since

$$y = \frac{2\pi}{\lambda} (x - a),$$

therefore we have uniformly along the whole number line

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos \frac{2\pi k}{\lambda} (x - a) + b_k \sin \frac{2\pi k}{\lambda} (x - a) \right]; \quad (1)$$

applying to the expressions

$$\cos \frac{2\pi k}{\lambda} (x - a), \quad \sin \frac{2\pi k}{\lambda} (x - a)$$

the trigonometrical formulae for the cosine and sine of difference of two arguments we evidently obtain the expansion (1) in the form

$$f(x) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \left(\alpha_k \cos \frac{2\pi k}{\lambda} x + \beta_k \sin \frac{2\pi k}{\lambda} x \right),$$

where α_k, β_k are constants; in particular, when $a \leq x \leq a + \lambda$, the sum of this series coincides with the corresponding values of the function $f(x)$ given in that interval.

We can thus see that the function $f(x)$ defined in an arbitrary interval $(a, a + \lambda)$ can be expanded into a generalised uniformly convergent trigonometrical series in that interval (provided the above conditions are satisfied) which differs from the series 1 § 81 only in that instead of the functions of the system (2) § 81 we have, in this case, the following elements of expansion

$$1, \cos \frac{2\pi n}{\lambda} x, \quad \sin \frac{2\pi n}{\lambda} x \quad (n = 1, 2, \dots),$$

which, as can be readily shown, form an orthogonal system in the interval $(a, a + \lambda)$ (and in every interval of length λ). All functions of this system have a period λ . As before, we find the following expressions for the coefficients α_k, β_k

$$\alpha_k = \frac{2}{\lambda} \int_a^{a+\lambda} f(x) \cos \frac{2\pi k}{\lambda} x dx \quad (k = 0, 1, 2 \dots),$$

$$\beta_k = \frac{2}{\lambda} \int_a^{a+\lambda} f(x) \sin \frac{2\pi k}{\lambda} x dx \quad (k = 1, 2, \dots).$$

Finally, if the conditions $f(a + \lambda) = f(a)$, $f'(a + \lambda) = f'(a)$ which are necessary for this expansion are not satisfied, we can always continue the function $f(x)$ beyond the point $a + \lambda$ to a certain point $b > a + \lambda$ so that all the requirements are satisfied in the interval (a, b) and we can then expand the function $f(x)$ in the extended interval in the way described above.

For exercises to § 85 *cf.* Problem Book by B.P. Demidovich, Nos. 331, 356, 357.

CHAPTER XXII

DIFFERENTIATION OF FUNCTIONS OF SEVERAL VARIABLES

§ 86. Continuity of functions of several independent variables

When we first introduced the concept of functional dependence (chapter I), we regarded the function $y = f(x)$ of one independent variable x as the simplest form of this general concept. The types of functional dependence which we meet in practice usually involve quantities which depend not on one but several (sometimes very many) other quantities whose values can be chosen arbitrarily and independent of one another so that we can call them independent variables. Only when these values are selected in a definite way, the function acquires a definite value. However, we have so far only studied the simplest case, *viz.* of one independent variable. In fact, the methods of differential and integral calculus can be successfully used for functions of any number of independent variables.

In this chapter we will, as a rule, consider in detail only the extension of methods and concepts of differential and integral calculus to functions of two independent variables and we shall leave the reader to prove for himself that increase in the number of variables does not involve any other new ideas. As we shall see, this method is useful insofar as the transition from functions of one independent variable to functions of two variables introduces some entirely new points and, in order to study them fully, we shall have to concentrate on this problem without dissipating our attention on a too bulky formal apparatus. On the other hand, the transition from two to three or more independent variables involves, as a rule, only some technical difficulties which can readily be overcome after the theoretical part has been fully understood.

Let u be a function of two independent variables x and y . In the same way as we have so far represented different values of the variable x by points on a straight line ("number line") and called these values "points" in some cases, we shall from now on often regard the independent variables x and y as the rectilinear coordinates of the point in a plane which we shall call the "number plane"; now, instead of speaking about "a pair of values (x, y) of the independent variables" we shall simply speak of the "point (x, y) " and the value of u for $x = a, y = b$ as its value at the "point (a, b) ". This terminology is very convenient in two respects: firstly it is in most cases much shorter (while giving the same degree of accuracy) and, secondly, it naturally tends to give a visual representation which in many cases makes it easier to understand the problem in question. In some cases it is convenient to denote the point (x, y) by a single letter, for example P, Q, \dots and write in short $u = f(P)$ instead of $u = f(x, y)$; this method is not often used in dealing with two independent variables since this abbreviated notation is not very useful; however, when we have several independent variables, this abbreviation can be very useful.

In the same way as the function $y = f(x)$ is represented geometrically in the rectilinear system of coordinates (x, y) by a *curve* which is only intersected at a single point by an arbitrary straight line parallel to the OY -axis, so the function $u = f(x, y)$ can be represented in the rectilinear system of coordinates (x, y, u) by a *surface* which is only intersected at a single point by an arbitrary straight line parallel to the OU -axis (this is equivalent to the condition that not more than one value of u should correspond to each pair of values of the variables x, y). We know the significance of graphical representations in the study of functions of one variable x . The geometrical illustration of functions of two variables by means of a surface in the three-dimensional space is equally important in the study of their theory.

Let the function $u = f(x, y)$ be defined in region ^{*)} of the plane and let $P(a, b)$ be a point in this plane. Apart from this point, let us consider another point $P'(x, y)$ which we shall imagine to move

*) At present the form of this region is irrelevant: it may be a square, a circle or any other finite figure, or the infinite part of the "number surface" or the whole of this surface.

indefinitely close to P so that $x \rightarrow a$ and $y \rightarrow b$; this can be expressed by the relation,

$$\rho = \sqrt{(x-a)^2 + (y-b)^2} \rightarrow 0.$$

If in this process the quantity $f(x, y)$ tends to a definite limit A , we can write

$$f(x, y) \rightarrow A \quad (\rho \rightarrow 0) \quad \text{or} \quad \lim_{\rho \rightarrow 0} f(x, y) = A.$$

It is very easy to describe the process by means of the relation $\rho \rightarrow 0$; however, we must note that in this case we are dealing with the limiting process in its most general form as considered in detail in § 15; in fact, $f(x, y)$ is not a function of the "basic variable" ρ , for evidently at a given distance ρ from P an infinite number of points $P'(x, y)$ lie where the function $f(x, y)$ will in general have different values. Nevertheless, the above relation has a quite definite meaning: no matter how small $\epsilon > 0$ be, there exists a $\delta > 0$ such that for any point $P'(x, y)$, other than P , which is at a distance not less than δ (i.e. the only condition is that $\rho < \delta$) we have:

$$|f(x, y) - A| < \epsilon.$$

Having thus defined the concept of limit for a function of two variables it is obvious to define the concept of continuity for such functions which we have found to be of fundamental importance in the case of functions of one variable.

The function $u = f(x, y)$ is said to be continuous at the point (x, y) if, assuming that $\rho = \sqrt{\Delta x^2 + \Delta y^2}$, we have:

$$\lim_{\rho \rightarrow 0} f(x + \Delta x, y + \Delta y) = f(x, y).$$

In detail this implies as follows: there exists a $\delta > 0$ for any $\epsilon > 0$ such that, provided $\rho < \delta$, we have:

$$|\Delta u| = |f(x + \Delta x, y + \Delta y) - f(x, y)| < \epsilon.$$

We can see that as before, the concept of continuity is of *local* character: in general the function of two variables can be continuous at some points and discontinuous at others. And again, as before, we shall agree to call the function $f(x, y)$ continuous within the given region of the number plane if it is continuous at every point in this plane.

While studying the functions of one independent variable we have considered in detail the structure of the continuum (the set of real numbers), *i.e.* the set of values of the independent variable itself, before studying the main properties of continuous functions (chapter 5). Similarly in this case too we must begin with the detailed study of the properties of sets of *pairs* (x, y) of real numbers before studying the properties of functions of two variables. This set is usually called the *two-dimensional continuum*; geometrically it is represented by a plane.

We have so far spoken of functions of *two* independent variables; however, all that is said in this paragraph also holds for functions of any number of variables so that we only need to give brief introduction.

If u is a function of n independent variables, it is convenient to call the set of values of these functions a *point in an n -dimensional space* (or a *space of n dimensions*). The *distance* between two points is equal to the square root of the sum of the squares of differences of the corresponding coordinates at these points (for $n = 3$ we have the usual distance between two points in the three-dimensional space). The function $u = f(x, y, z)$ is *continuous* at the point (x, y, z) of the three-dimensional space if

$$\Delta u = f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z) \rightarrow 0 \quad (\rho \rightarrow 0),$$

where

$$\rho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

(there is no need to give the definition of continuity for $n > 3$, as it is self-evident). In order to study the further development of the science of functions of n variables it is necessary to consider in detail the properties of an n -dimensional continuum, *i.e.* the set of all groups of n real numbers (a_1, a_2, \dots, a_n) .

The student is advised to consider the instructive examples in the Problem Book by B.P. Demidovich, Section VI, Nos. 44, 45, 55.

§ 87. Two-dimensional continuum

On a straight line we can have only one type of simple figure, *viz.* an interval. However, in the transition to a two-dimensional continuum, *viz.* the plane, we can have a great variety of simple figures: polygons, circles and, in general, other figures bounded by simple contours. The variety of these figures introduces many new points

in the study of a two-dimensional continuum as compared with the simplest linear (one-dimensional) continuum.

The set of all points of this simple figure is called the *region* which is closed if all points on the contour belong to the region and *open* if no point on the contour belongs to the region. For the time being, we shall disregard regions extending to infinity : thus, for example, we shall call a region (closed) semi-plane $x > 0$ or the whole number plane in the same way as we have earlier regarded the semi-straight line $x > 0$ or even the whole number line as being particular cases of intervals.

If the point $P(x, y)$ is an interior point (*i.e.* it does not lie on the contour) of the region D , then any sufficiently small circle with centre at P together with all its points will belong to the whole region. On the other hand, if P is a point on the contour (or, as it is usually said, on the *boundary*) of the region D , then any circle with centre at P will contain points which belong to the region D as well as other points which do not belong to it. These properties can be regarded as definitions of interior and boundary points of the region. An open region consists solely of interior points ; a closed region contains, apart from interior points, all boundary points too.

In the case of the linear continuum the dimension of simplest figures are fully described by their lengths. In the case of planes the position is somewhat more complicated : in relation to the problem in consideration we may be interested in the area of the given region or in its linear dimensions ; the latter are best described by the *diameter* of the region which is defined as *the upper bound of the common distances between all possible pairs of points belonging to the given region* ; thus the diameter of a circle is the length of its usual diameter, the diameter of a rectangle is the length of its diagonal, and so on. If the upper bound, which we have mentioned above, exists, the region is said to be bounded ; otherwise it is *infinite* ; an infinite region is sometimes said to have a diameter equal to $+\infty$. In order that a region should be bounded it is evidently necessary and sufficient that it should lie entirely within a circle (in the same way as a linear set is bounded if and only if it lies entirely within an interval).

Let us assume that we are given a region D and a point P in a plane and let $\rho(P, Q)$ denote the distance between two points P and Q in the plane. If Q runs through different points of the region D , then $\rho(P, Q)$ has a definite lower bound which we shall call the *distance of the point P from the region D* and denote by $\rho(P, D)$.

Theorem 1. *If the point P does not belong to the closed region D , then $\varphi(P, D) > 0$.*

Proof. Where $\varphi(P, D) = 0$, then any arbitrarily small circle with centre at P would contain points of the region D . In that case there are two possibilities :

(1) any sufficiently small circle with centre at P belongs entirely to the region D , or

(2) any circle with centre at P contains points of the region D and other points which do not belong to that region.

In the first case the point P would, according to the definition, be an interior point and in the second case it would be a boundary point of the region D . Since the region D is closed, the point P would in either case belong to that region, which contradicts the conditions of the theorem. Hence $\varphi(P, D) > 0$ and theorem 1 is proved.

Analogous to the theorem on a contracting sequence of sections (§ 18, lemma 1) we must now establish the corresponding important theorem on contracting sequences of closed regions. The sequence of regions $D_1, D_2, \dots, D_n, \dots$ with corresponding diameters $d_1, d_2, \dots, d_n, \dots$ is said to be *contracting* if 1) $D_{n+1} \subset D_n$ ($n = 1, 2, \dots$) (i.e. the region D_{n+1} lies entirely within the region D_n , and 2) $d_n \rightarrow 0$ ($n \rightarrow \infty$).

Theorem 2. *The contracting sequence of closed regions always has one common point for all regions of the given sequence.*

Proof. Let us denote by (a_n, b_n) the interval which is projection of the region D_n on the OX -axis and by (c_n, d_n) a similar interval on the OY -axis. Evidently each of the two sequences (a_n, b_n) ($n = 1, 2, \dots$) and (c_n, d_n) ($n = 1, 2, \dots$) represents a contracting sequence of intervals. Let α be the common point for all intervals (a_n, b_n) (it follows from lemma 1 § 18 that such a point exists and is unique) and let β be the common point for all intervals (c_n, d_n) . We say that point $P(\alpha, \beta)$ belongs to each of the regions D_n . In fact, if a closed region D_k exists which does not contain the point P , then, according to theorem 1, we would have $\varphi(P, D_k) = d > 0$. But $D_l \subset D_k$ for $l > k$, and therefore

$$\varphi(P, D_l) \geq \varphi(P, D_k) = d \quad (l \geq k). \quad (1)$$

But if $Q(x, y)$ is an arbitrary point of the region D_l , then x and α belong to the interval (a_l, b_l) and β and y to the interval (c_l, d_l) so that

$$\varphi(P, Q) = \sqrt{(x - \alpha)^2 + (y - \beta)^2} \leq \sqrt{(b_l - a_l)^2 + (d_l - c_l)^2},$$

and therefore

$$\rho(P, D_l) \leq \sqrt{(b_l - a_l)^2 + (d_l - c_l)^2}; \quad (2)$$

but $b_l - a_l \rightarrow 0$ and $d_l - c_l \rightarrow 0$ for $l \rightarrow \infty$ and therefore, according to (2), $\rho(P, D_l) \rightarrow 0$, which contradicts the inequality (1) according to which $\rho(P, D_l) \geq d$ for every $l \geq k$. Hence the point $P(\alpha, \beta)$ belongs to each region D_n . If another point P' , possessing the same property, exists, then let us assume that the distance between the points P and P' is equal to ρ ; in this case each of the regions D_n evidently contains both points P and P' and its diameter should not be less than ρ . But this contradicts the condition that the region D_n forms a contracting sequence. This proves uniqueness of the point P .

We shall now prove the "theorem on finite coverage" which is analogous to lemma 2 §18. Let us assume that we are given (a finite or infinite) set (system) S of regions (D). We shall say that the system S covers a certain region Δ if each point of the region Δ is an interior point of at least one of the regions D of the system S .

Theorem 3. *If the system S covers the bounded closed region Δ , then another system can be separated from this system, which consists of a finite number of regions and which also covers the region Δ .*

Proof. Since the region Δ is bounded, it lies entirely within a square Q . Let us divide this square into four equal squares by drawing straight lines through the middle points of its opposite sides. We shall say that a square is "normal" if the part of the region Δ which it contains does not permit finite coverage required by theorem 3 (the square which contains no points of the region Δ is not regarded normal). Theorem 3 evidently implies that the square Q is not normal. On the contrary let us assume that it is normal; it can then readily be seen that out of the four squares into which it has been divided at least one square must be normal; in fact, if each of these squares would permit finite coverage (or if it would contain no points of the region Δ), then, evidently, the square Q as a whole would also permit finite coverage.

Let Q_1 be a normal quarter; dividing it again into four squares we see that at least one of these four squares must be normal and so on. We can continue this process as long as we please and as a result obtain a contracting sequence of squares Q, Q_1, Q_2, \dots . Let P be the common point of these squares (according to theorem 2 this

point exists and is unique). Let us at first show that P belongs to the region Δ . In fact, any circle with centre at P evidently contains all the squares Q_n provided n is sufficiently large and therefore it also contains points of the region Δ . If, when the radius is sufficiently small, this circle entirely lies within the region Δ , the point P is an interior point of this region; if however, when the radius is as small as we please, this circle contains both the points belonging to Δ and the point not belonging to this region, the point P lies on the boundary of the region Δ ; and since the region Δ is closed, the point P belongs to this region in either case.

It therefore follows from the theorem that a region D of the system S exists where P is an interior point. Any sufficiently small circle with centre at P therefore entirely belongs to the region D ; but such a circle contains, as we know, an infinite number of squares Q_n , each of which is thus covered by *one* region D of the system S whereas, according to its definition, it is normal and cannot permit finite coverage. This contradiction shows that our assumption cannot be correct, *i.e.*, the square Q cannot be normal and therefore it should permit finite coverage. Theorem 3 is thus proved.

Let us now assume that D_1 and D_2 are two bounded closed regions which have no points in common and let P be an arbitrary point of the region D_1 . According to theorem 1 there exists a circle with centre at P which contains no points of the region D_2 . Let $r(P)$ be the radius of this circle. Let us agree to call a circle of radius $\frac{1}{2}r(P)$ with centre at P as "proper" circle of the point P , and let us denote by S the set of "proper" circles of all points P of the region D_1 . Since the system S covers the region D_1 , therefore, according to theorem 3, there exists a finite group S' of circles of the group S , which also covers the region D_1 ; let us denote by δ the radius of the smallest circle of this finite group S' .

Let P_1 and P_2 be two arbitrary points which belong to the regions D_1 and D_2 respectively. The point P_1 lies inside a circle belonging to the group S' ; let P and $r = \frac{1}{2}r(P)$ denote respectively the centre and radius of this circle. We then have firstly $\rho(P, P_2) > r(P)$ (since the point P_2 belongs to the region D_2) and secondly $\rho(P, P_1) < r = \frac{1}{2}r(P)$. Therefore

$$\rho(P_1, P_2) \geq \rho(P, P_2) - \rho(P, P_1) > r(P) - \frac{1}{2}r(P) = \frac{1}{2}r(P) \geq \delta.$$

And since P_1 is an arbitrary point of the region D_1 and P_2 an arbitrary point of the region D_2 therefore we can conclude that the

lower bound of all the distances $\rho(P_1, P_2)$ is a *positive* number. This lower bound is called the *common distance between the regions* D_1 and D_2 and denoted by $\rho(D_1, D_2)$. We thus arrive at the following important proposition.

Theorem 4. *If D_1 and D_2 are bounded closed regions with no points in common, then $\rho(D_1, D_2) > 0$.*

All the concepts, statements and proofs of theorems given in this paragraph can be extended to a continuum of arbitrary dimensions without making any essential changes, as the reader can readily show himself.

§ 88. Properties of continuous functions

We now possess sufficient knowledge in order to establish the main properties of continuous functions of several variables.

We at first note that all theorems which we have proved in §§ 21 and 22 for functions of one variable also hold for functions of two variables. As a result of rational operations with functions continuous at an arbitrary point P we again obtain a function continuous at that point (in the case of division it is only necessary that divider should not vanish at the point P). The theorem on continuity of a composite function should, in this case, be understood in the following sense: if $z = f(u, v)$, $u = \varphi_1(x, y)$, $v = \varphi_2(x, y)$ and if the functions φ_1 and φ_2 are continuous at the point $P(x, y)$ in the XY -plane whereas the function $f(u, v)$ is continuous at the points $u = \varphi_1(x, y)$, $v = \varphi_2(x, y)$ of the UV -plane, then the function

$$F(x, y) = f[\varphi_1(x, y), \varphi_2(x, y)],$$

which, as given in the above form, is called a composite function of x and y , is continuous at the point P .

All these theorems are proved in exactly the same way as the analogous theorems in § 21 and § 22 and we therefore leave the proof to the reader.

We shall now enumerate several more important properties of functions of two variables.

Theorem 1. *The function $f(x, y)$ continuous in the closed bounded region D is bounded in that region.*

The proof of this theorem is so similar to the proof of the analogous theorem 1 § 23 that the reader will have no difficulties in proving it by himself.

Theorem 2. *The function $f(x, y)$ continuous in the bounded region D takes its maximum and minimum values in that region.*

We can again leave the proof to the reader, for it is analogous to the proof of the corresponding theorem 2 § 23.

The concept of *uniform continuity* for functions of two variables is in all respects similar to that of functions of one variable and it is equally important. The function $f(x, y)$ is said to be *uniformly continuous* in the region D if the following condition holds: no matter how small $\varepsilon > 0$ be, there exists a $\delta > 0$ such that for any two points P_1 and P_2 of the region D which are at a distance

$$\rho(P_1, P_2) < \delta$$

we have:

$$|f(P_1) - f(P_2)| < \varepsilon.$$

The following theorem which is similar to theorem 5 § 23 also holds:

Theorem 3. *The function $f(x, y)$ continuous at every point of the bounded closed region D is uniformly continuous in that region.*

Although the proof of this theorem is quite similar to that of theorem 5 § 23 we shall nevertheless give it here in full since it is rather complicated.

Proof. Let P be an arbitrary point of the region D and ε an arbitrary positive number. Since the function $f(x, y)$ is continuous at the point P , therefore at any point P' of a circle with centre at P and a sufficiently small radius $\rho(P)$ we have:

$$|f(P') - f(P)| < \frac{\varepsilon}{2};$$

therefore if P' and P'' are two arbitrary points of the above circle (bearing in mind that P' and P'' belong to the region D) we have:

$$|f(P') - f(P'')| < \varepsilon. \quad (1)$$

We can construct a similar circle for every point P of the above region D ; the radii $\rho(P)$ of these circles will, of course, differ from

one another. Let us now imagine that every point P of the region D is surrounded not only by our constructed circle of radius $\rho(P)$ but also by another concentric circle with radius equal to half of the radius of the constructed circle, *viz.* $\frac{1}{2}\rho(P)$. This circle is called the *proper circle* of the point P .

Owing to the fact that each point P of the region D has a proper circle, therefore the set S of all proper circles will cover the region D . It therefore follows from theorem 3 of the previous paragraph that there should exist a finite group C_1, C_2, \dots, C_n of proper circles which would also cover the region D . Each of these circles C_k has its centre at P_k and radius equal to $\frac{1}{2}\rho(P_k)$. Let us denote by δ the smallest of these n radii.

Let P' and P'' be two arbitrary points of the region D which are at a distance

$$\rho(P', P'') < \delta,$$

and let the point P' belong to the circle C_k with centre at P_k and radius $\frac{1}{2}\rho(P_k)$. We then have

$$\rho(P', P_k) \leq \frac{1}{2}\rho(P_k), \quad (2)$$

whereas for $\delta \leq \frac{1}{2}\rho(P_k)$,

$$\rho(P', P'') < \frac{1}{2}\rho(P_k); \quad (3)$$

it follows from (2) and (3) that

$$\rho(P'', P_k) < \rho(P_k),$$

i.e. the point P'' belongs to the circle with centre at P_k and radius $\rho(P_k)$; and since the point P' also belongs to that circle therefore the inequality (1) holds. But ϵ is arbitrarily small and P' and P'' are two arbitrary points of the region D which are at a distance less than δ . Therefore the function $f(x, y)$ is uniformly continuous in the region D , which was to be proved.

We leave the reader to show that all results and proofs given in this paragraph can be extended without essential modifications to continuous functions of any number of variables.

§ 89. Partial derivatives

We shall now study the theory of differential calculus relating to functions of several variables and at first deal with functions of

two variables. Here again the evaluation of the rate of change of the function is very important; however, in this case the problem is much more complicated. Earlier the course of the process was described by the variation of one variable x and all we had to do was to observe the rate of change of the function $y = f(x)$ when the variable received this or other increment; on the other hand, we are now dealing with a point $P(x, y)$ in a plane; this point can not only be displaced in distance but also in *any direction* and it is clear that in general the rate of change of the function $u = f(x, y)$ will differ when this point is displaced in different directions. Hence in order to solve this problem fully we shall have to analyse all aspects of this complicated picture.

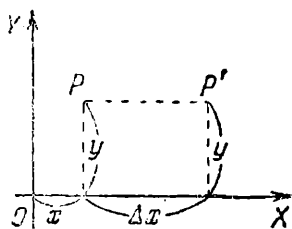


Fig. 52.

However, we shall at first only consider the simplest case when out of the two coordinates x and y of the point P only one changes while the other remains constant so that the point P is displaced in the direction of one of the coordinate axes. Let, for example, the variable x receive an increment Δx while the variable y remains constant so that we go from the point $P(x, y)$ to the point $P'(x + \Delta x, y)$ (fig. 52). The function $u = f(x, y)$ evidently receives in this process the following increment

$$\Delta u = f(x + \Delta x, y) - f(x, y).$$

The ratio $\Delta u / \Delta x$ thus gives us the *average rate of change of the function u with respect to the variable x in the interval $(x, x + \Delta x)$ for the given constant value of the variable y* . If the following limit

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta u}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

exists for $\Delta x \rightarrow 0$, we can evidently regard it as the *local rate of change of the function $u = f(x, y)$ with respect to the variable x at the point $P(x, y)$* . It is obvious that x and y are constant in this limiting process, while Δx alone changes. However, the magnitude of the limit will evidently depend on the chosen values of x and y ; generally speaking it will be different at different points (x, y) and, like u , it is a function of x and y (that is why we speak of the *local rate*). This quantity is known as *partial derivative* of the function $u = f(x, y)$ with respect to x and denoted by $\partial u / \partial x$ or $\partial f(x, y) / \partial x$

or $f'_x(x, y)$. In the first case the letter ∂ (round) and in the second case the lower index x show that we are differentiating with respect to *one* of the two variables while the value of the other remains unchanged (fixed).

Thus

$$\frac{\partial u}{\partial x} = \frac{\partial f(x, y)}{\partial x} = f'_x(x, y) = \lim_{\Delta x \rightarrow 0} \frac{\Delta u}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}.$$

Similarly the following quantity

$$\frac{\partial u}{\partial y} = \frac{\partial f(x, y)}{\partial y} = f'_y(x, y) = \lim_{\Delta y \rightarrow 0} \frac{\Delta u}{\Delta y} = \lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y},$$

is known as partial derivative of the function $u = f(x, y)$ and expresses the local rate of change of this function with respect to the variable y ; here the rate of change of u is involved when the point $P(x, y)$ is displaced in the direction of the OY -axis. Hence by knowing both partial derivatives $\partial u / \partial x$ and $\partial u / \partial y$ at the point P we are evidently able to judge the rate of change of the function $u = f(x, y)$ when the point $P(x, y)$ is displaced in the direction of one of the coordinate axes; however, from partial derivatives of a function we cannot judge about its rate of change when the point P is displaced in other directions.

From what has been said above it is clear that finding of partial derivatives of functions given in concrete form does not necessitate the use of any new methods. Thus, in order to find, say, $\partial u / \partial x$ it is sufficient to find the usual derivative of the function $u = f(x, y)$ with respect to x assuming y to be constant in this process, so that u becomes a function of one variable x .

Example. $u = y^2 \sin 3x, \quad \frac{\partial u}{\partial x} = 3y^2 \cos 3x$

$$\frac{\partial u}{\partial y} = 2y \sin 3x.$$

Partial derivatives of a function of two variables can be readily illustrated geometrically. The equation $u = f(x, y)$ expresses a surface in three-dimensional space; by giving y an arbitrary fixed value, for example $y = b$, we concentrate our attention on a cross-section of

this surface by the plane $y = b$; this cross-section is a flat curve whose equation has the form

$$u = f(x, b) \quad (1)$$

(fig. 53). The partial derivative $\partial u / \partial x$ at an arbitrary point $P(a, b)$ is the usual derivative of the function $f(x, b)$ with respect to x at the point a and is therefore equal to the angular coefficient of the tangent to the curve (1) drawn at the point $x = a$ (i. e. the tangent of the angle between the direction of the tangent and the positive direction of the OX -axis). $\partial u / \partial y$ can be illustrated in exactly the same way.

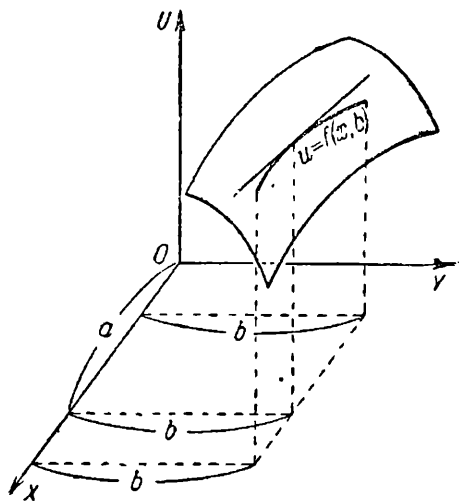


Fig. 53

Partial derivatives of functions of three or more independent variables are determined similarly; if, for example,

$$u = f(x, y, z),$$

then

$$\frac{\partial u}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y, z) - f(x, y, z)}{\Delta x};$$

$\partial u / \partial y$ and $\partial u / \partial z$ are determined similarly. It is obvious that a function of n independent variables has n different partial derivatives. Each of these derivatives expresses a local (calculated at the given point of the n -dimensional space) rate of change of the function in the direction of the corresponding axis of coordinates. This geometrical representation can only be imagined visually for $n = 1, 2$ and 3.

Not many exercises need to be done in connection with this paragraph. The student should only solve 5–6 problems chosen from among Nos. 66–81, Section VI of the Problem Book by B. P. Demidovich.

§ 90. Differentials

When we were studying the function $y = f(x)$ of one variable x , each displacement of the point x to an arbitrary new point $x + \Delta x$ corresponded to another quantity dy which is called differential

of the function y corresponding to the given displacement of the point x . The differential dy is defined as the product $f'(x) \Delta x$ and has the following two properties: (1) it is proportional to the displacement Δx of the point x and (2) when $\Delta x \rightarrow 0$, it differs from the increment $\Delta y = f(x + \Delta x) - f(x)$ of the function y by an infinitely small quantity of a higher order as compared to Δx . Many applications of the differential are based on these two properties. We have also learnt that these two properties fully define the differential of the function y so that no other dy possessing the same properties can exist.

While dealing with a function $u = f(x, y)$ of two variables it is desirable to construct a quantity possessing analogous properties. Let us assume that we go from the point $P(x, y)$ to another point $P'(x + \Delta x, y + \Delta y)$. Let us denote by $\rho = \sqrt{\Delta x^2 + \Delta y^2}$ the distance between these two points. We wish to construct a quantity which would play a similar part for the function $u = f(x, y)$ as the differential for functions of one variable and correspond to the transition (displacement) of the point P to the point P' . Since the differential of one variable has the form $A \Delta x$, where A is independent of Δx , (but generally it depends on x) therefore, in this case, the quantity du should be a linear combination of the increments Δx and Δy , i. e. it should have the form

$$du = A \Delta x + B \Delta y, \quad (1)$$

where A and B are independent of Δx and Δy (but they generally depend on x and y). This condition evidently corresponds to the first of the above properties of a differential. Its second property involves the fact that the difference between the increment and the differential is an infinitely small quantity of a higher order as compared to the magnitude of the displacement which in the first case is measured by $|\Delta x|$ and here by the distance $\rho = \sqrt{\Delta x^2 + \Delta y^2}$ between the points P and P' . Therefore in the transition from the point P to the point P' it is necessary that the differential du of the function $u = f(x, y)$ should differ from its increment $\Delta u = f(x + \Delta x, y + \Delta y) - f(x, y)$ by an infinitely small quantity of a higher order as compared to ρ .

Both these conditions can be combined if we accept the following definition.

The differential du of the function $u = f(x, y)$ at the point $P(x, y)$

is an expression of the form (1), where A and B are independent of Δx and Δy if for $\rho = \sqrt{\Delta x^2 + \Delta y^2} \rightarrow 0$

$$\Delta u - du = o(\rho) \quad (2)$$

where Δu is the increment of the function u received in transition from the point $P(x, y)$ to the point $P'(x + \Delta x, y + \Delta y)$.

The coefficients A and B cannot be determined directly from this definition and we can draw no conclusions as to existence and uniqueness of the differential of the given function $u = f(x, y)$. We shall now consider these problems.

Theorem 1. *If the function $u = f(x, y)$ has the differential (1) at the point $P(x, y)$, then both partial differentials $\partial u / \partial x$ and $\partial u / \partial y$ also exist at that point and $A = \partial u / \partial x$, $B = \partial u / \partial y$, so that*

$$du = \frac{\partial u}{\partial x} \Delta x + \frac{\partial u}{\partial y} \Delta y.$$

Proof. Let x receive the increment Δx while y remains constant ($\Delta y = 0$) i.e. we displace the point P in the direction of the OX -axis; we thus have $du = A \Delta x$ and therefore

$$\Delta u = du + o(\rho) = A \Delta x + o(|\Delta x|),$$

Since in our case we evidently have $\rho = |\Delta x|$. Therefore

$$\frac{\Delta u}{\Delta x} = \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} = A + \frac{o(|\Delta x|)}{\Delta x} = A + o(1)$$

when $\Delta x \rightarrow 0$, the limit of the right-hand side is equal to A ; hence the left-hand side also has a limit which is also equal to A ; in other words, $\partial u / \partial x$ exists and is equal to A ; we can prove similarly that $\partial u / \partial y$ exists and is equal to B . Hence theorem 1 is proved.

Uniqueness of the differential follows from theorem 1.

Theorem 1 shows that, as in the case of functions of one variable, existence of a differential implies existence of the partial derivatives $\partial u / \partial x$ and $\partial u / \partial y$. In the case of functions of one variable the converse proposition is also true: existence of a derivative implies existence of a differential so that we can define differentiability of a function either by existence of a derivative or of a differential, since these two definitions are equivalent; it is therefore natural to ask a similar question in the case of functions of two variables: let it be

given that the function $u = f(x, y)$ has both partial derivatives $\partial u / \partial x$ and $\partial u / \partial y$ at the point P ; does this fact imply existence of the differential du at that point? It is not difficult to foresee that this condition is not compulsory in this case; with functions of one variable the derivative fully describes the rate of change of the function; in the case under consideration the knowledge of $\partial u / \partial x$ and $\partial u / \partial y$ only makes it possible to determine this rate in two out of an infinite number of possible directions; hence it only describes variation of the given function to a very limited extent.

It can be readily shown that, in general, existence of partial derivatives at a given point does not, in fact, guarantee existence of a differential. Let us consider the function $u = f(x, y) = \sqrt{|xy|}$ in the neighbourhood of the point $(0, 0)$. Since $u = 0$ everywhere along the OX and OY axes, therefore $\partial u / \partial x = 0$ and $\partial u / \partial y = 0$ at the point $(0, 0)$. If the function u would have had a differential at that point, then, according to theorem 1, we should have for every Δx and Δy

$$du = 0;$$

it therefore follows from (2) that $\Delta u = o(\rho)$. But if we choose $\Delta x = \Delta y > 0$, we have:

$$\rho = \sqrt{\Delta x^2 + \Delta y^2} = \Delta x \sqrt{2}, \quad \Delta u = f(\Delta x, \Delta x) - f(0, 0) = \Delta x,$$

and when $\Delta x \rightarrow 0$, the increment Δu is of the same order of smallness as compared to the displacement ρ . We thus arrive at a contradiction which shows that the function u does not have a differential at the point $(0, 0)$ although both its partial derivatives exist at that point.

Therefore in the case of functions of two variables existence of a differential is a stronger condition than that of partial derivatives. However, we must regard cases when there exist partial derivatives but no differentials as exceptions rather than the general rule. This is shown by the following theorem which assures existence of differentials in many real cases.

Theorem 2. *If at the point (x, y) the partial derivatives $\partial u / \partial x$ and $\partial u / \partial y$ of the function $u = f(x, y)$ exist and are continuous, then the function has a differential at that point.*

Obviously the condition that the functions $\partial u / \partial x$ and $\partial u / \partial y$ are continuous implies that these functions also exist in a neighbour-

hood of the point (x, y) (in a circle with centre at (x, y) and a sufficiently small radius), for otherwise continuity of these functions at the point (x, y) would be void.

Proof. Assuming that $\rho = \sqrt{\Delta x^2 + \Delta y^2}$, $du = \partial u / \partial x \cdot \Delta x + \partial u / \partial y \cdot \Delta y$ we must prove that for $\rho \rightarrow 0$

$$\Delta u - du = o(\rho).$$

We have:

$$\begin{aligned} \Delta u &= f(x + \Delta x, y + \Delta y) - f(x, y) = \\ &= f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) + f(x, y + \Delta y) - f(x, y). \end{aligned} \quad (3)$$

The right-hand side of this equation contains the sum of two differences. Let us consider the first of these. The second argument of both terms has the same value $y + \Delta y$; we can therefore regard this difference as the increment Δx of a function of one variable x . If $|\Delta x|$ and $|\Delta y|$ are sufficiently small, then this function has a derivative at every point of the interval $(x, x + \Delta x)$; this derivative is nothing but the partial derivative of the function u with respect to the variable x in the immediate neighbourhood of the point (x, y) whose existence is preassumed. We can therefore apply the theorem on finite increments (§ 36) and write

$$f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) = f'_x(x + \theta_1 \Delta x, y + \Delta y) \Delta x,$$

where $0 < \theta_1 < 1$. Similarly we obtain for the second difference on the right-hand side of the equation (3)

$$f(x, y + \Delta y) - f(x, y) = f'_y(x, y + \theta_2 \Delta y) \Delta y,$$

where $0 < \theta_2 < 1$. Hence equation (3) gives:

$$\Delta u = f'_x(x + \theta_1 \Delta x, y + \Delta y) \Delta x + f'_y(x, y + \theta_2 \Delta y) \Delta y,$$

and consequently

$$\begin{aligned} \Delta u - du &= [f'_x(x + \theta_1 \Delta x, y + \Delta y) - f'_x(x, y)] \Delta x + \\ &\quad + [f'_y(x, y + \theta_2 \Delta y) - f'_y(x, y)] \Delta y, \end{aligned}$$

and since evidently

$$|\Delta x| \leq \rho, \quad |\Delta y| \leq \rho,$$

therefore

$$\begin{aligned} \frac{|\Delta u - du|}{\rho} &\leq |f'_x(x + \theta_1 \Delta x, y + \Delta y) - f'_x(x, y)| + \\ &\quad + |f'_y(x, y + \theta_2 \Delta y) - f'_y(x, y)|. \end{aligned}$$

Since the functions $f'_x(x, y)$ and $f'_y(x, y)$ are continuous at the point (x, y) , therefore both terms on the right-hand side tend to zero for $\rho \rightarrow 0$; we therefore have

$$\lim_{\rho \rightarrow 0} \frac{\Delta u - du}{\rho} = 0$$

or, which is the same,

$$\Delta u = du + o(\rho);$$

this proves that $du = \partial u / \partial x \cdot \Delta x + \partial u / \partial y \cdot \Delta y$ is the differential of the function u at the point (x, y) .

Let us also make the following remark. If we have

$$u = f(x, y) = x,$$

then

$$\frac{\partial u}{\partial x} = 1, \quad \frac{\partial u}{\partial y} = 0,$$

and consequently

$$du = dx = \Delta x;$$

similarly by considering the function $u = y$ we arrive at the conclusion that $dy = \Delta y$. Hence in this case, as in earlier cases, the increments of the independent variables and differentials are equivalent to one another. This also shows that the differential of the function u , in case it exists, can be written in the form

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy.$$

Owing to the fact that with functions of two variables existence of a differential is, in general, not equivalent to existence of partial derivatives, the question naturally arises, when the function of two variables is differentiable at a given point. The answer to this question is, to a certain extent, provided by what we have learnt so far. As we know, the differential describes the behaviour of the function during displacement in any direction whereas partial derivatives only tell us as to what happens when the point is displaced in two well defined (mutually perpendicular) directions (the partial derivatives may even disappear when the coordinate axes are rotated about the given point). It is therefore natural to say that the given function $f(x, y)$ is *differentiable* only at points where it has a *differential* and we cannot be satisfied by the mere existence

of partial derivatives. The usefulness of this definition of differentiability will be confirmed on many future occasions and particularly in the next § 91.

The concept of differential and all its properties can be extended without essential modifications to functions of three or more variables. Thus in the case of the function $u = f(x, y, z)$ we define the differential du at the point (x, y, z) by the following expression

$$du = A \Delta x + B \Delta y + C \Delta z,$$

where A , B and C are independent of Δx , Δy and Δz provided $\Delta u - du = o(\rho)$ when $\rho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \rightarrow 0$. As before, we can readily show that for existence of the differential it is necessary that $A = \partial u / \partial x$, $B = \partial u / \partial y$, $C = \partial u / \partial z$. In particular, $dx = \Delta x$, $dy = \Delta y$, $dz = \Delta z$, so that

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy + \frac{\partial u}{\partial z} dz.$$

The function u which has a differential at the point (x, y, z) is said to be differentiable at that point. Existence of partial derivatives of the function u at the point (x, y, z) does not guarantee its differentiability at that point; however, if all three partial derivatives at the point (x, y, z) are *continuous*, then the function u is differentiable at that point.

For exercises to § 90, cf. Problem by B.P. Demidovich, Section VI, Nos. 96a, 97a, 104.

§ 91. Derivatives in arbitrary directions

We have already said on several occasions that partial derivatives of functions of two variables only define the rate of change in the direction of the axes of coordinates; there are no general reasons for isolating these two directions from among all other possible directions; we must therefore now consider the rate of change of the function $u = f(x, y)$ when the point (x, y) is displaced in an arbitrary direction.

Let us draw a straight line through the point $P(x, y)$ to make an arbitrary

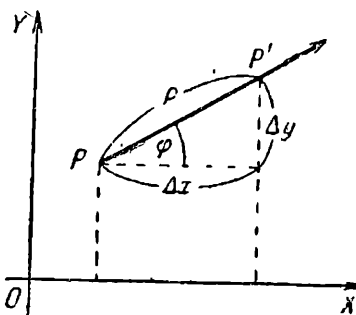


Fig. 54

angle φ with the positive direction of the OX -axis (fig. 54) and move from the point $P(x, y)$ to an arbitrary point $P'(x + \Delta x, y + \Delta y)$ which lies on the drawn straight line. The distance ρ between the points P and P' is evidently equal to $\sqrt{\Delta x^2 + \Delta y^2}$. Dividing the increment $\Delta u = f(x + \Delta x, y + \Delta y) - f(x, y)$ received by the given function in the transition from the point P to the point P' by the distance ρ between these points we obtain a quantity which can be regarded as the average rate of change of the function in this transition; if this average rate tends to a definite limit for $\rho \rightarrow 0$, then this limit should be regarded as the (local) rate of change of the function at the given point in the direction determined by that of the drawn straight line which, for the sake of brevity is called as the "direction φ ". This limit, in case it exists, is called the *derivative in the direction φ* of the function $u = f(x, y)$ at the given point (x, y) and denoted by $D_\varphi f(x, y)$ or $D_\varphi u$.

Hence

$$D_\varphi f(x, y) = \lim_{\rho \rightarrow 0} \frac{f(x + \Delta x, y + \Delta y) - f(x, y)}{\rho},$$

where $\rho = \sqrt{\Delta x^2 + \Delta y^2}$ and where it is assumed that Δx and Δy change in such a way that the point $(x + \Delta x, y + \Delta y)$ always lies on the straight line drawn in the direction φ (for this to be so it is evidently necessary and sufficient that we should always have $\Delta y/\Delta x = \tan \varphi$). It is clear that the partial derivative $\partial u/\partial x$ is the common value of the quantities $D_{\partial/\partial x} u$ and $-D_{\pi/\partial x} u$ when these two quantities coincide; similarly $\partial u/\partial y$ is the common value of the coinciding quantities $D_{\partial/\partial y} u$ and $-D_{3\pi/2/\partial y} u$.

Let us now assume that the function $u = f(x, y)$ is differentiable at the point (x, y) , i.e. that it has a differential which, as we know, is equal to

$$du = \frac{\partial u}{\partial x} \Delta x + \frac{\partial u}{\partial y} \Delta y,$$

and in the transition from the P to the point P'

$$\Delta u = f(x + \Delta x, y + \Delta y) - f(x, y) = du + o(\rho),$$

hence

$$\frac{\Delta u}{\rho} = \frac{\partial u}{\partial x} \frac{\Delta x}{\rho} + \frac{\partial u}{\partial y} \frac{\Delta y}{\rho} + \frac{o(\rho)}{\rho},$$

or, since $\Delta x = \rho \cos \varphi$, $\Delta y = \rho \sin \varphi$ (cf. fig. 54), therefore

$$\frac{\Delta u}{\rho} = \frac{\partial u}{\partial x} \cos \varphi + \frac{\partial u}{\partial y} \sin \varphi + \frac{o(\rho)}{\rho}.$$

When the point P' approaches indefinitely close to the point P in the direction φ , i.e. when $\rho \rightarrow 0$, the first two terms on the right-hand side of the above equation remain constant while the third term tends to zero. We therefore obtain in the limit :

$$D_{\varphi}u = \lim_{\rho \rightarrow 0} \frac{\Delta u}{\rho} = \frac{\partial u}{\partial x} \cos \varphi + \frac{\partial u}{\partial y} \sin \varphi.$$

We have thus proved the following proposition which evidently confirms the usefulness of the chosen definition for differentiability of functions of two variables.

Theorem 1. *If the function $u = f(x, y)$ is differentiable at the point P , then it has derivatives in all directions φ at that point and*

$$D_{\varphi}u = \frac{\partial u}{\partial x} \cos \varphi + \frac{\partial u}{\partial y} \sin \varphi.$$

We can thus see that for a function to be differentiable it is necessary to know the partial derivatives (i.e. the derivatives in two mutually perpendicular directions); this enables us to write down the expression for the derivative in any desired direction without performing additional calculations.

When the given direction is changed to the directly opposite direction (i.e. from φ to $\varphi + \pi$), both $\cos \varphi$ and $\sin \varphi$ change their signs; we therefore have for any direction φ :

$$D_{\varphi+\pi}(u) = -D_{\varphi}(u),$$

i.e. the absolute values of the derivatives in two mutually opposite directions are the same but have opposite signs. We already saw that this holds for derivatives in the direction of the coordinate axes.

Let us now consider the same problem for a function of three independent variables $u = f(x, y, z)$. Let us draw a straight line through the point $P(x, y, z)$ to make angles α , β and γ respectively with the coordinate axes and take an arbitrary point $P'(x + \Delta x, y + \Delta y, z + \Delta z)$ on that line. Let us denote by

$$\rho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

the distance between the points P and P' ; we evidently have :

$$\Delta x = \rho \cos \alpha, \Delta y = \rho \cos \beta, \Delta z = \rho \cos \gamma.$$

Let us now assume that the function u is differentiable at the point (x, y, z) so that for $\rho \rightarrow 0$

$$\begin{aligned} \Delta u &= f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z) = \\ &= \frac{\partial f}{\partial x} \rho \cos \alpha + \frac{\partial f}{\partial y} \rho \cos \beta + \frac{\partial f}{\partial z} \rho \cos \gamma + o(\rho). \end{aligned}$$

If the point P' comes indefinitely close to the point P while always remaining on the drawn straight line, we shall have $\rho \rightarrow 0$, while the angles α, β and γ remain constant; we therefore have :

$$\lim_{\rho \rightarrow 0} \frac{\Delta u}{\rho} = \frac{\partial f}{\partial x} \cos \alpha + \frac{\partial f}{\partial y} \cos \beta + \frac{\partial f}{\partial z} \cos \gamma.$$

This limit, as in the case of functions of two variables, evidently describes the local (at the point $P(x, y, z)$) rate of change of the function u during the displacement of the point P in the chosen direction which is characterised by the angles α, β and γ . Therefore we can in this case also say that the above limit is the derivative Du of the function u in the given direction. Thus *if the function $u = f(x, y, z)$ is differentiable at the point P , then it has a derivative Du at that point in any direction (α, β, γ) and*

$$Du = \frac{\partial u}{\partial x} \cos \alpha + \frac{\partial u}{\partial y} \cos \beta + \frac{\partial u}{\partial z} \cos \gamma. \quad (1)$$

For exercises to § 91 cf. Problem Book by B.P. Demidovich, Section VI, Nos. 198, 199, 201.

§ 92. Differentiation of composite and implicit functions

We shall now consider two problems dealing with the usual differentiation of functions of one variable which we could not study at the time, since their solution requires knowledge of the methods for differentiating functions of several variables.

1. Let $u = f(x, y)$ be a function of two variables x and y where each is regarded not as an independent variable but as a function of a new variable t : $x = \varphi(t)$, $y = \psi(t)$ (which is the same in both cases). We thus have

$$u = f[\varphi(t), \psi(t)]$$

which is now composite function of t . We must find the derivative du/dt of u with respect to the independent variable t from the partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ and the derivatives $dx/dt = \varphi'(t)$ and $dy/dt = \psi'(t)$. We are given that all these derivatives exist but existence of the required derivative du/dt must be proved. We shall assume with regard to the function $f(x, y)$ that it not only has partial derivatives but also differential at the point $x = \varphi(t)$, $y = \psi(t)$.

Let the increment Δt of t correspond to the increments Δx and Δy of x and y which in their turn correspond to the increment Δu of u . Let us assume that $\sqrt{\Delta x^2 + \Delta y^2} = \rho$. We know that

$$\Delta u = du + o(\rho) = \frac{\partial u}{\partial x} \Delta x + \frac{\partial u}{\partial y} \Delta y + \alpha \rho$$

where $\alpha \rightarrow 0$ for $\rho \rightarrow 0$. Therefore

$$\frac{\Delta u}{\Delta t} = \frac{\partial u}{\partial x} \frac{\Delta x}{\Delta t} + \frac{\partial u}{\partial y} \frac{\Delta y}{\Delta t} \pm \alpha \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2}. \quad (1)$$

Let us now assume that $\Delta t \rightarrow 0$. Therefore $\Delta x \rightarrow 0$ and $\Delta y \rightarrow 0$ (since x and y are differentiable, they are also continuous); hence $\rho \rightarrow 0$ and also $\alpha \rightarrow 0$. But since, on the other hand, the derivatives $\partial u/\partial x$ and $\partial u/\partial y$ are constant for $\Delta t \rightarrow 0$ and the ratios $\Delta x/\Delta t$ and $\Delta y/\Delta t$ tend to the limits $dx/dt = \varphi'(t)$ and $dy/dt = \psi'(t)$ respectively, therefore the right-hand side of the relation (1) has the following limit for $\Delta t \rightarrow 0$:

$$\frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt}.$$

This proves that $\frac{du}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta u}{\Delta t}$ also exists and

$$\frac{du}{dt} = \frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt}. \quad (2)$$

This simple formula evidently solves our problem.

The following case which often appears in applications deserves special attention, *i. e.* when $x = \varphi(t) = t$, *i. e.* when the new variable t is the same as one of the old variables. This means that u is given as a function of two variables x and y , where x is the independent variable and $y = \psi(x)$ is a given function of x : $u = f[x, \psi(x)]$. Assuming in formula (2) that $dt = dx$ we obtain:

$$\frac{du}{dx} = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \frac{dy}{dx}. \quad (3)$$

The left-hand side and the first term on the right-hand side of this formula represent the derivatives of the function u with respect to the independent variable x ; however, these derivatives do not coincide with one another, for they are found on different bases: $\partial u/\partial x$ is the *partial derivative* of u with respect to x , *i. e.* it is calculated on the assumption that y remains constant; on the other hand du/dx is a "full" derivative of u with respect to x , *i. e.* it is evaluated on the assumption that $y = \psi(x)$ changes in a quite definite manner as x changes. Formula (3) clearly shows the importance of different symbols used for denoting partial and full derivatives.

Example. $u = \frac{y}{x}, y = \sqrt{1-x^2}$;

$$\frac{\partial u}{\partial x} = -\frac{y}{x^2}, \frac{\partial u}{\partial y} = \frac{1}{x}; \quad \frac{dy}{dx} = -\frac{x}{\sqrt{1-x^2}} = -\frac{x}{y};$$

according to formula (3) we have :

$$\frac{du}{dx} = -\frac{y}{x^2} - \frac{1}{x} \cdot \frac{x}{y} = -\frac{x^2 + y^2}{x^2 y} = -\frac{1}{x^2 \sqrt{1-x^2}}.$$

The problem of the derivative du/dt when u is given as the differential of a function of any number of variables, each of which, in its turn, is a differentiable function of t , is solved in exactly the same way. Thus when $u = f(x, y, z)$, $x = \varphi(t)$, $y = \psi(t)$, $z = \chi(t)$, we have :

$$\frac{du}{dt} = \frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt} + \frac{\partial u}{\partial z} \frac{dz}{dt} = \frac{\partial u}{\partial x} \varphi'(t) + \frac{\partial u}{\partial y} \psi'(t) + \frac{\partial u}{\partial z} \chi'(t). \quad (4)$$

The reader will have no difficulty in deducing this formula in the same way as formula (2) is deduced above.

In all the cases which we have considered so far the problem involves *one* independent variable (which is denoted by t). However, it often happens in applications that there are several independent variables. Let us assume that we have, as above, $u = f(x, y)$ but let this time x and y be functions of two variables $x = \varphi(t, s)$, $y = \psi(t, s)$. In this case $u = f[\varphi(t, s), \psi(t, s)]$ also becomes a function of the same variables; we thus have a composite function. Our problem involves the following: we know the partial derivatives of the function $f(x, y)$ with respect to x and y and also the partial derivatives of the functions $\varphi(t, s)$ and $\psi(t, s)$ with respect to t and s ; we are required to find the partial derivatives of u with respect to t and s . Since partial differentiation does not differ in

any way from ordinary differentiation (but is carried out under definite conditions), this problem does not require further consideration. It is solved by means of relations analogous to formula (2) :

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial t},$$

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial s};$$

we obtain similar results when the number of intermediate functions or the number of independent variables is greater than two.

2. Let us now consider the second problem. We are given an arbitrary equation with two variables

$$F(x, y) = 0. \quad (5)$$

Generally speaking this equation defines one or several functions y of the independent variable x *); thus, for example, the equation

$$xy - 2x + 3y - 1 = 0$$

defines one function $y = (2x+1)/(x+3)$ and the equation $x^2 + y^2 - 1 = 0$ two functions: $y = +\sqrt{1-x^2}$ and $y = -\sqrt{1-x^2}$ **); many cases are known when functions defined by the given equation (5) cannot be expressed in terms of x by means of elementary formulae in the way it is done in the above two simple examples; however, regardless of whether such an expression is possible, every function $y = f(x)$ which within a region of values of x identically satisfies the

*) The conditions under which this takes place will be considered in detail in chapter 24.

**) Strictly speaking we only obtain two functions when we restrict ourselves to *continuous* solutions of the given equation. The above equation defines an infinite number of *discontinuous* functions. Any one of the following functions is a solution of this equation :

$$y = \begin{cases} -\sqrt{1-x^2} & (-1 \leq x \leq \alpha), \\ +\sqrt{1-x^2} & (\alpha \leq x \leq 1), \end{cases}$$

where α is an arbitrary number between -1 and $+1$. The most general solution of the above equation can be written as follows: $y = \psi(x) \sqrt{1-x^2}$, where $\sqrt{1-x^2} \geq 0$ and $\psi(x)$ is an arbitrary function which only takes the values $+1$ and -1 .

equation (5) is said to be an *implicit* function described by this equation. It is our problem to find derivative of such a function.

Let $y = f(x)$ be an implicit function given by the equation (5); we therefore have in a region

$$F[x, f(x)] = 0,$$

for every x and therefore in this region we also have

$$\frac{dF[x, f(x)]}{dx} = 0$$

for every x . Let us now assume that both the functions $F(x, y)$ and $f(x)$ are differentiable; we then obtain from formula (3) :

$$\frac{dF[x, f(x)]}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \frac{dy}{dx};$$

hence when $y = f(x)$ we have in that region

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \frac{dy}{dx} = 0,$$

and therefore

$$\frac{dy}{dx} = - \frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} \quad (6)$$

(provided, of course, that $\partial F / \partial y \neq 0$).

This formula solves our problem. However, we must make the following remark. It may appear strange that we have succeeded in finding a “definite” expression (6) for the derivative dy/dx though we were unable to express the function itself in a “definite” form, *i.e.* to solve the equation (5) with respect to y . The solution of this equation would require that y should be expressed in terms of x by means of an elementary formula; it would indeed be strange if we were unable to do this for the function y , while we could always do it for its derivative dy/dx ; however, formula (6) does not give us such an expression for dy/dx ; since $\partial F / \partial x$ and $\partial F / \partial y$ are given as functions of two variables x and y , formula (6) expresses the derivative dy/dx in terms of these two variables. Hence if we do not know a definite expression for the function y in terms of x , then formula (6) cannot give us either a definite expression in terms of x for the derivative of this function.

Nevertheless, formula (6) which establishes a connection between the derivative of an implicit function and the partial derivatives of a function of two variables which describe it is of great importance in theory and has many important applications, some of which we shall consider later.

Example. Let y be an implicit function of x given by the equation

$$F(x, y) = xy^5 - x^5y - 2 = 0;$$

we have:

$$\frac{\partial F}{\partial x} = y^5 - 5x^4y, \quad \frac{\partial F}{\partial y} = 5xy^4 - x^5,$$

and formula (6) gives:

$$\frac{dy}{dx} = -\frac{y(y^4 - 5x^4)}{x(5y^4 - x^4)}.$$

For exercises to § 92 cf. Problem Book by B. P. Demidovich, Section VI, Nos. 137-139, 223-226, 229, 231.

§ 93. Homogeneous functions and Euler theorem

A polynomial $P(x, y)$ of two variables is said to be homogeneous if sum of the indices of the variables x and y of all its terms has the same value k ; the latter is known as *degree of homogeneity* of the polynomial. Thus the polynomial

$$ax^3 + bx^2y + cxy^2 + dy^3$$

will be homogeneous of degree 3 regardless of the coefficients a, b, c and d .

If $P(x, y)$ is a homogeneous polynomial of degree k , we evidently have for every t (and for all x, y)

$$P(tx, ty) = t^k P(x, y).$$

This property of homogeneous polynomials is very useful for the development of the concept of homogeneous functions. We shall agree to say that a *homogeneous function of degree k* is a function $f(x, y)$ which identically satisfies the following relation (*i e.* for all values of x, y, t):

$$f(tx, ty) = t^k f(x, y). \quad (1)$$

In contrast to polynomials the index k can, in this case, have any real value; it is self-evident that in such cases t can only take values for which t^k has a definite meaning; thus, for example, we should have $t \neq 0$ for $k < 0$, $t \geq 0$ for $k = \frac{1}{2}$, etc.

Examples. $(x^2 + y^2)/(x + y)$, $(x - y)/(x + y)$, $(x + y)/(x^2 + y^2)$ are homogeneous functions of degrees respectively equal to 1, 0 and -1 .

The following simple and very convenient relation developed by Euler holds for partial derivatives of homogeneous functions. Let $f(x, y)$ be a homogeneous function of degree k . We shall assume that x and y are constant in the relation (1) while t is variable so that both sides of this relation are functions of t . We can then differentiate the identity (1) with respect to t . On the left-hand side we evidently have a composite function of t whose derivative can be found by means of formula (2) § 92: assuming that $tx = u$, $ty = v$, we obtain:

$$\frac{df(u, v)}{dt} = \frac{\partial f(u, v)}{\partial u} \frac{du}{dt} + \frac{\partial f(u, v)}{\partial v} \frac{dv}{dt} = x \frac{\partial f(u, v)}{\partial u} + y \frac{\partial f(u, v)}{\partial v}.$$

The derivative of the right-hand side of the relation (1) is equal to $kt^{k-1}f(x, y)$; the two derivatives obtained are equal to one another for all values of x, y, t ; assuming, in particular, that $t = 1$, we obtain $u = x, v = y$ and therefore

$$x \frac{\partial f(x, y)}{\partial x} + y \frac{\partial f(x, y)}{\partial y} = kf(x, y).$$

This is Euler's relation. The same method can be readily used for finding analogous relations for homogeneous functions of any number of variables: we shall only state the result for functions of three variables.

The function $f(x, y, z)$ is said to be a *homogeneous function of degree k* if the following relation holds identically:

$$f(tx, ty, tz) = t^k f(x, y, z);$$

if this function is differentiable, we have:

$$x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} + z \frac{\partial f}{\partial z} = kf.$$

For exercises cf. Problem Book by B.P. Demidovich. Section VI, No. 93.

§ 94. Partial derivatives of higher orders

The partial derivatives $\partial u / \partial x$ and $\partial u / \partial y$ of the function $u = f(x, y)$ are functions of the same variables x and y on which u depends. Therefore the operations of partial differentiation with respect to either of these variables can again be applied to these functions. The partial derivatives with respect to x and y of the functions $\partial u / \partial x$ and $\partial u / \partial y$ are said to be *derivatives of the second order* with regard to the initial function u . Each of the derivatives $\partial u / \partial x$ and $\partial u / \partial y$ gives rise to two derivatives of the second order so that we obtain in total four derivatives of the second order usually denoted as follows :

$$\frac{\partial}{\partial x} \left(\frac{\partial u}{\partial x} \right) = \frac{\partial^2 u}{\partial x^2} = f''_{xx}(x, y),$$

$$\frac{\partial}{\partial y} \left(\frac{\partial u}{\partial x} \right) = \frac{\partial^2 u}{\partial x \partial y} = f''_{xy}(x, y),$$

$$\frac{\partial}{\partial x} \left(\frac{\partial u}{\partial y} \right) = \frac{\partial^2 u}{\partial y \partial x} = f''_{yx}(x, y),$$

$$\frac{\partial}{\partial y} \left(\frac{\partial u}{\partial y} \right) = \frac{\partial^2 u}{\partial y^2} = f''_{yy}(x, y).$$

Each of these four derivatives of the second order is a function of the same variables x and y and can in its turn have partial derivatives with respect to these variables which we shall call derivatives of the third order of the function u ; derivatives of the third order are defined in exactly the same way as above so that no additional explanations are needed; thus

$$\frac{\partial^3 u}{\partial x^2 \partial y} = f'''_{xxy}(x, y)$$

denotes a function obtained as a result of differentiating the function $u = f(x, y)$ three times where the first two differentiations are carried out with respect to x and the third differentiation with respect to y . In general the partial derivatives of the first order of any derivative of order n are said to be partial derivatives of order $n + 1$ of the function u and denoted in the way described above. It is evident that for a function of two variables the number of derivatives of the third order is equal to eight and, in general, the number of derivatives of order n is equal to 2^n . Partial derivatives of higher orders are very important in mathematics of accurate nature study; they are also extensively used in mathematical physics.

Partial derivatives of higher orders possess one very important property which considerably simplifies the set of these derivatives and their formulae. This property is due to the fact that if two partial derivatives of the same order differ from one another only by the order in which differentiation is performed and if they are both continuous, then they are exactly alike.

Let us at first consider derivatives of the second order. The functions $f''_{xy}(x, y)$ and $f''_{yx}(x, y)$ are obtained from the function $u = f(x, y)$ as a result of two differentiations. In both cases one of these differentiations is with respect to x and the other with respect to y ; the only difference is due to the order in which these differentiations are performed. We say that if at a certain point (x, y) the functions f''_{xy} and f''_{yx} are both continuous, then

$$f''_{xy}(x, y) = f''_{yx}(x, y).$$

In order to prove this let us consider the expression

$$\Delta = f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y) - f(x, y + \Delta y) + f(x, y).$$

Assuming, when y and Δy are constant, that

$$f(x, y + \Delta y) - f(x, y) = \varphi(x)$$

we can evidently write

$$\Delta = \varphi(x + \Delta x) - \varphi(x). \quad (1)$$

Existence of the second derivatives of the function f implies existence of the first derivatives in the neighbourhood of the point (x, y) ; therefore the function $\varphi(x)$ is differentiable in the interval $(x, x + \Delta x)$, provided Δx and Δy are sufficiently small. Applying the theorem on finite increments to the right-hand side of formula (1) we obtain :

$$\Delta = \varphi'(x + \theta_1 \Delta x) \Delta x,$$

where $0 < \theta_1 < 1$. But it follows from the definition of the function $\varphi(x)$ that

$$\varphi'(x) = f'_x(x, y + \Delta y) - f'_x(x, y),$$

so that we obtain :

$$\Delta = [f'_x(x + \theta_1 \Delta x, y + \Delta y) - f'_x(x + \theta_1 \Delta x, y)] \Delta x. \quad (2)$$

But, on the other hand, the application of the theorem on finite increments to the difference in the square brackets on the right-hand side of the above equation evidently gives :

$$\begin{aligned} f'_x(x + \theta_1 \Delta x, y + \Delta y) - f'_x(x + \theta_1 \Delta x, y) &= \\ &= f''_{xy}(x + \theta_1 \Delta x, y + \theta_2 \Delta y) \Delta y, \end{aligned}$$

where again $0 < \theta_2 < 1$. Therefore the relation (2) gives :

$$\Delta = f''_{xy}(x + \theta_1 \Delta x, y + \theta_2 \Delta y) \Delta x \Delta y. \quad (3)$$

Let us now return to the initial expression for the quantity Δ and transform it in another way. Let us assume that

$$f(x + \Delta x, y) - f(x, y) = \psi(y),$$

so that

$$\Delta = \psi(y + \Delta y) - \psi(y),$$

or, applying the theorem on finite increments,

$$\Delta = \psi'(y + \theta_3 \Delta y) \Delta y,$$

where $0 < \theta_3 < 1$. It follows from the definition of the function $\psi(y)$ that

$$\psi'(y) = f'_y(x + \Delta x, y) - f'_y(x, y),$$

and we obtain :

$$\Delta = [f'_y(x + \Delta x, y + \theta_3 \Delta y) - f'_y(x, y + \theta_3 \Delta y)] \Delta y,$$

or, applying again the theorem on finite increments

$$\Delta = f''_{yx}(x + \theta_4 \Delta x, y + \theta_3 \Delta y) \Delta x \Delta y \quad (0 < \theta_4 < 1). \quad (4)$$

Comparing the equations (3) and (4) we obtain for $\Delta x \Delta y \neq 0$

$$f''_{xy}(x + \theta_1 \Delta x, y + \theta_2 \Delta y) = f''_{yx}(x + \theta_4 \Delta x, y + \theta_3 \Delta y).$$

Let us now assume that Δx and Δy tend to zero so that we always have $\Delta x \Delta y \neq 0$. Since, according to our assumption, the functions f''_{xy} and f''_{yx} are continuous at the point (x, y) , therefore the limiting process gives :

$$f''_{xy}(x, y) = f''_{yx}(x, y), \quad (5)$$

which was to be proved.

Let us now consider the general case. Let us assume that we have at first two partial derivatives of the same order $n \geq 2$ which

differ from one another only in that in the first of these derivatives two consecutive differentiations are carried out in an order reverse of the other, for example

$$f^{(5)}_{xyxy}, \quad f^{(5)}_{yxxy},$$

where the difference is due to the commutability of the second and third differentiations and all other operations are carried out in exactly the same order in both cases. However, it follows from the above proof that two such derivatives are exactly alike (in view of the necessary condition of continuity); thus, by applying equation (5) to the function $f'_x(x, y)$ in our example, we obtain :

$$f'''_{xy} = f'''_{yx};$$

continuing differentiation on both sides of this equation first with respect to x and then with respect to y , we obtain an equation for the two given derivatives of the fifth order.

However, in the most general case when we are given two arbitrary derivatives of order n which differ from one another only by the order of differentiations, we can evidently pass from one to the other by exchanging two successive differentiations. Since these commutative operations leave the derivative unchanged, the two derivatives will remain alike.

The proved proposition considerably reduces the number of different derivatives of order n and gives a better insight into the set of these derivatives. In fact, if the order of differentiations is irrelevant, then we can evidently obtain any desired derivative by differentiating first with respect to x and then with respect to y for the required number of times; therefore any derivative of order n of the function u can be represented in the form

$$\frac{\partial^n u}{\partial x^k \partial y^{n-k}},$$

where k is one of the numbers of the series $0, 1, \dots, n$. This shows directly that the number of different derivatives of order n is equal to $n + 1$ whereas we had 2^n such derivatives earlier, *i.e.* when n was large, the number of derivatives was many times greater.

The definitions and notations used for partial derivatives of higher orders also hold for functions depending on three or more variables. The possibility of changing the order of differentiations

also holds for such functions provided the functions which are compared with one another are continuous. The proof of this theorem follows directly from what is said above, since all changes in the order of the differentiations for functions of any number of variables can evidently be brought about by a series of commutative operations of two successive differentiations and such a change, as we have shown, leaves the result unaltered.

For exercises to §94, cf. Problem Book by B.P. Demidovich. Section VI, Nos. 82, 112, 113, 118-120, 162, 164, 166.

§ 95. Taylor's formula for functions of two variables

All considerations which at the time (chapter 9) prompted us to represent functions of one variable by Taylor's formula remain fully valid for functions of any number of variables: here, as before, it is very convenient both theoretically and practically to represent the given function approximately in the form of a polynomial of a given degree. On the other hand, our initial assumptions as to the validity of this representation are the same as before. At that time we obtained Taylor's formula by developing the simple formula

$$f(x+h) - f(x) = hf'(x) + o(h),$$

which holds for every differentiable function. For a differentiable function of two variables $u = f(x, y)$ we have an analogous formula

$$\Delta u = f(x+h, y+k) - f(x, y) = hf'_x(x, y) + kf'_y(x, y) + o(\rho),$$

where $\rho = \sqrt{\Delta x^2 + \Delta y^2}$. We have therefore good reasons for trying to obtain in this case an approximate expression for the quantity $f(x+h, y+k)$ in the form of a polynomial in powers of the increments h and k . Taylor's formula could, in fact, be obtained by repeating with corresponding changes and complications the whole deduction which in chapter 9 gave us Taylor's formula for functions of one variable.

However, there is a much simpler and shorter way which will give us the desired result if, instead of starting from the very beginning, we assume that we have already established Taylor's formula for functions of one variable. This method is also convenient in that it is carried out in exactly the same way for functions of any number of variables and for the sake of brevity we only restrict ourselves here to the consideration of functions of two

variables. We shall assume that the values x and y and their increments h and k are constant and we shall consider the function

$$\varphi(t) = f(x + ht, y + kt) \quad (1)$$

of one variable t in the interval $0 \leq t \leq 1$. Let us assume that the function $f(x, y)$ has all partial derivatives inclusive upto the order n and all these derivatives are differentiable at the point (x, y) . It then follows from formula (2) § 92 that the derivative $\varphi'(t)$ exists and is equal to

$$\varphi'(t) = \frac{\partial f}{\partial x} \frac{d(x + ht)}{dt} + \frac{\partial f}{\partial y} \frac{d(y + kt)}{dt} = h \frac{\partial f}{\partial x} + k \frac{\partial f}{\partial y},$$

where both partial derivatives are taken at the point $(x + ht, y + kt)$. Applying the same formula to the function $\varphi'(t)$ and using the fact that as a result of (5) § 94

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x},$$

we readily obtain:

$$\varphi''(t) = h^2 \frac{\partial^2 f}{\partial x^2} + 2hk \frac{\partial^2 f}{\partial x \partial y} + k^2 \frac{\partial^2 f}{\partial y^2},$$

where again all partial derivatives are taken at the point $(x + ht, y + kt)$. Using the same method we obtain further:

$$\varphi'''(t) = h^3 \frac{\partial^3 f}{\partial x^3} + 3h^2k \frac{\partial^3 f}{\partial x^2 \partial y} + 3hk^2 \frac{\partial^3 f}{\partial x \partial y^2} + k^3 \frac{\partial^3 f}{\partial y^3}.$$

These formulae show that in case all constituent partial derivatives exist and are differentiable, the following general formula holds:

$$\varphi^{(n)}(t) = \sum_{r=0}^n C_n^r h^{n-r} k^r \frac{\partial^n f}{\partial x^{n-r} \partial y^r}, \quad (2)$$

where all partial derivatives are taken at the point $(x + ht, y + kt)$. Formula (2) proved for $n = 1, 2, 3$ can be in general proved by means of induction from n to $n + 1$, as we shall now see; this proof is simple and clear in principle but it involves rather bulky calculations.

Let us assume that formula (2) is valid for a given n and let all its constituent partial derivatives be differentiable at the point

$(x + ht, y + kt)$. Applying formula (2) § 92 to the function $\varphi^n(t)$ we obtain :

$$\begin{aligned}\varphi^{(n+1)}(t) &= \sum_{r=0}^n C_n^r h^{n-r} k^r \left\{ h \frac{\partial^{n+1} f}{\partial x^{n-r+1} \partial y^r} + k \frac{\partial^{n+1} f}{\partial x^{n-r} \partial y^{r+1}} \right\} = \\ &= \sum_{r=0}^n C_n^r h^{n+1-r} k^r \frac{\partial^{n+1} f}{\partial x^{n+1-r} \partial y^r} + \sum_{r=0}^n C_n^r h^{n-r} k^{r+1} \frac{\partial^{n+1} f}{\partial x^{n-r} \partial y^{r+1}} = \Sigma_1 + \Sigma_2.\end{aligned}$$

In the second of these sums we shall change the index of summation by assuming that $r = s - 1$; hence

$$\Sigma_2 = \sum_{s=1}^{n+1} C_n^{s-1} h^{n+1-s} k^s \frac{\partial^{n+1} f}{\partial x^{n+1-s} \partial y^s},$$

or, denoting the new index of summation by r as before

$$\Sigma_2 = \sum_{r=1}^{n+1} C_n^{r-1} h^{n+1-r} k^r \frac{\partial^{n+1} f}{\partial x^{n+1-r} \partial y^r}. \quad (3)$$

Let us also note the following fact: if we agree to consider for every n

$$C_n^{-1} = C_n^{n+1} = 0,$$

then we can sum from $r = 0$ to $r = n + 1$ the sum Σ_1 and the expression (3) for Σ_2 without altering these sums in any way. We therefore obtain :

$$\varphi^{(n+1)}(t) = \sum_{r=0}^{n+1} (C_n^r + C_n^{r-1}) h^{n+1-r} k^r \frac{\partial^{n+1} f}{\partial x^{n+1-r} \partial y^r}.$$

But according to a well-known property of binomial coefficients

$$C_n^r + C_n^{r-1} = C_{n+1}^r \quad (0 \leq r \leq n + 1);$$

and we therefore obtain :

$$\varphi^{(n+1)}(t) = \sum_{r=0}^{n+1} C_{n+1}^r h^{n+1-r} k^r \frac{\partial^{n+1} f}{\partial x^{n+1-r} \partial y^r},$$

i.e. formula (2) remains valid when n is replaced by $n + 1$; this proves the formula in its most general form.

It also establishes existence of a derivative of order n for the function $\varphi(t)$ if, as assumed, the function $f(x, y)$ has differentiable partial derivatives inclusive up to that order. It therefore follows that MacLaurin formula holds for $\varphi(t)$:

$$\begin{aligned}\varphi(t) &= \varphi(0) + t\varphi'(0) + \frac{t^2}{2!}\varphi''(0) + \dots \\ &\dots + \frac{t^{n-1}}{(n-1)!}\varphi^{(n-1)}(0) + \frac{t^n}{n!}\varphi^{(n)}(\theta t),\end{aligned}$$

where the last term is written in Lagrange's special form which is well-known to us from § 39; in particular for $t = 1$

$$\begin{aligned}\varphi(1) &= \varphi(0) + \varphi'(0) + \frac{1}{2!}\varphi''(0) + \dots \\ &\dots + \frac{1}{(n-1)!}\varphi^{(n-1)}(0) + \frac{1}{n!}\varphi^{(n)}(\theta),\end{aligned}$$

where $0 < \theta < 1$. But $\varphi(0) = f(x, y)$, $\varphi(1) = f(x + h, y + k)$, and the successive derivatives of the function φ for $t = 0$ are expressed by formula (2) where all partial derivatives are taken at the point (x, y) . We therefore obtain:

$$\begin{aligned}f(x + h, y + k) &= f(x, y) + \left(h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y}\right) + \frac{1}{2!}\left(h^2\frac{\partial^2 f}{\partial x^2} + \right. \\ &+ 2hk\frac{\partial^2 f}{\partial x \partial y} + k^2\frac{\partial^2 f}{\partial y^2}\left.) + \dots + \frac{1}{(n-1)!}\sum_{r=0}^{n-1} C_{n-1}^r h^{n-1-r} k^r \frac{\partial^{n-1} f}{\partial x^{n-1-r} \partial y^r} + \right. \\ &\left. + \frac{1}{n!}\sum_{r=0}^n C_n^r h^{n-r} k^r \frac{\partial^n f(x + \theta h, y + \theta k)}{\partial x^{n-r} \partial y^r}, \quad (4)\end{aligned}$$

where $0 < \theta < 1$ and where all partial derivatives except those which enter the last sum (last term) on the right-hand side are taken at the point (x, y) (and are therefore independent of h and k).

The last formula completely solves our problem, for it gives an approximate expression for $f(x + h, y + k)$ in the form of a polynomial of degree $n - 1$ with respect to h and k and as

can be readily seen, the last term has the required form $o(\rho^{n-1})$ ($\rho = \sqrt{h^2 + k^2}$) so that $|h| \leq \rho$, $|k| \leq \rho$ and consequently

$$|h^{n-r} k^r| \leq \rho^n = o(\rho^{n-1}) \quad (0 \leq r \leq n).$$

The notation chosen for this formula is rather bulky (although the formula itself is clear and can be readily remembered in spite of its outward complexity). This bulkiness becomes even worse when we consider functions of three or more variables. Therefore Taylor's formula is often written in symbolic form. Let us write the following expression for an arbitrary natural number q :

$$\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}\right)^q f.$$

If we raise the binomial in the brackets to the q -th degree in accordance with the binomial formula and assume that ∂ is a number (and not the symbol of differentiation), we obtain:

$$\left[\sum_{r=0}^q C_q^r h^{q-r} k^r \frac{\partial^q}{\partial x^{q-r} \partial y^r}\right] f = \sum_{r=0}^q C_q^r h^{q-r} k^r \frac{\partial^q f}{\partial x^{q-r} \partial y^r},$$

i.e. (with an accuracy to the factor $1/q!$) the q -th degree in Taylor's formula. We can agree to write even more briefly

$$\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}\right)^q f = L_q f;$$

thus L_q becomes a certain definite operation to be performed over the function f which we have just described in full. By using this notation we can write Taylor's formula in symbolic form as follows:

$$\begin{aligned} f(x+h, y+k) &= \sum_{q=0}^{n-1} \frac{1}{q!} \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}\right)^q f(x, y) + \\ &+ \frac{1}{n!} \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}\right)^n f(x+\theta h, y+\theta k), \end{aligned}$$

or even more briefly

$$f(x+h, y+k) = \sum_{q=0}^{n-1} \frac{1}{q!} L_q f(x, y) + \frac{1}{n!} L_n f(x+\theta h, y+\theta k).$$

For exercises to § 95 *cf.* Problem Book by B.P. Demidovich, Section VI, Nos. 390, 391, 396, 398, 400, 406, 407.

§ 96. Extrema

The maxima and minima of a function of any number of variables in a given region of these variables are defined in exactly the same way as for functions of one variable. Here the concept of a *local extremum* is of fundamental importance: by this we mean an interior point in the given region at which the value of the function is not less (or not greater) than at any other point sufficiently close to the given point. As in the case of a function of one variable, a function can have an extremum either on the boundary of the region or at an interior point which will, in this case, also be a point of a local extremum. It is clear that in the space of several dimensions the problem is complicated by the fact that even in the simplest cases all boundary points, of which there is an infinite number, enter into competition (in the one-dimensional space the boundary consisted of only two points); we must therefore find the greatest or smallest value of the function on the boundary of the given region, *i.e.* we must solve an additional extremum problem. In practice some material considerations often make it possible to determine beforehand that the function attains, say, its maximum value inside (and not on the boundary of) the region and this considerably simplifies the solution of the problem. We must, nevertheless, use differential calculus in order to find the points of the local extremum.

If the function $u = f(x, y)$ has a local extremum at the point (a, b) , then it is clear that the function

$$\varphi(x) = f(x, b)$$

of one variable x should have a local extremum at the point $x = a$. Let us assume that the function $f(x, y)$ is differentiable everywhere in the given region; in that case the function $\varphi(x)$ evidently has a derivative equal to $f'_x(x, b)$ in a neighbourhood of the point a . We know from § 41 that therefore $\varphi'(a) = 0$, *i.e.* the partial derivative $\partial f / \partial x$ of the function $f(x, y)$ should vanish at the point of a local extremum $x = a, y = b$. Similarly, an analogous argument shows that we should also have $\partial f / \partial y = 0$ at that point. Finally it can be similarly shown that the result obtained remains valid for functions of any number of variables: *if such a function is differentiable in a region, then partial derivatives with respect to all the variables should be equal to zero at every point of a local extremum (which lies within the region).*

In this case a *stationary point* is a point at which the partial derivatives with respect to all the variables vanish; formula (1) § 91

shows that the derivative of the given function in every direction is equal to zero at a stationary point; hence a stationary point is, as it were, a point of minimum changeability of the function during a displacement in any direction; this justifies the use of this term.

Hence in the space of several dimensions, finding of extrema requires at first the knowledge of all stationary points of the given function in the given region. If we have a function of n variables, then by equating the partial derivatives of this function with respect to all the variables to zero, we obtain a system of n equations with n unknowns and thus determine the coordinates of the stationary points. This problem can now be solved without further using differential calculus.

Having found all stationary points we must, in the same way as in a one-dimensional case, investigate each point individually and determine whether it gives the maximum or minimum of the given function or neither. In the case of several dimensions this investigation is much more complicated and we shall here only give the first few steps for functions of two variables.

Let $P(a, b)$ be a stationary point of the function $u = f(x, y)$ and let this function have all partial derivatives of the second order at the point P . Let us also consider the point $Q(a + h, b + k)$ and denote the distance between these two points by ρ so that

$$\rho = \sqrt{h^2 + k^2};$$

finally, let A, B and C denote corresponding partial derivatives

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y^2}$$

at the point P . It follows from Taylor's formula ((4) § 95) that we have for $\rho \rightarrow 0^*$:

$$\Delta u = f(a + h, b + k) - f(a, b) = \frac{1}{2}(Ah^2 + 2Bhk + Ck^2) + o(\rho^2).$$

If we denote by α the angle made by the vector \overrightarrow{PQ} (the "displacement" of the point P) and the positive direction of the OX -axis, then evidently

$$h = \rho \cos \alpha, \quad k = \rho \sin \alpha,$$

*) Terms of the first order with respect to h and k vanish, since the point $P(a, b)$ is a stationary point.

and consequently

$$\begin{aligned}\Delta u &= f(\alpha + h, b + k) - f(\alpha, b) = \\ &= \frac{1}{2} \rho^2 (A \cos^2 \alpha + 2B \cos \alpha \sin \alpha + C \sin^2 \alpha) + o(\rho^2).\end{aligned}$$

Using this expression for the increment Δu we can readily see that the nature of the given stationary point $P(a, b)$ depends on the behaviour of the quantity

$$\varphi(\alpha) = A \cos^2 \alpha + 2B \cos \alpha \sin \alpha + C \sin^2 \alpha$$

which is a function of the “angle of displacement” and varies from 0 to 2π . Let, for example, $\varphi(\alpha) > 0$ ($0 \leq \alpha \leq 2\pi$). Since the function $\varphi(\alpha)$ is a continuous function of α , it assumes a smallest value in the interval $(0, 2\pi)$, which, according to our assumption, is positive; it follows from the expression obtained for Δu that we have for $\rho \rightarrow 0$

$$\Delta u = \rho^2 \left\{ \frac{1}{2} \varphi(\alpha) + o(1) \right\},$$

and since $\varphi(\alpha) \geq \mu$ ($0 \leq \alpha \leq 2\pi$), therefore for a sufficiently small ρ we have $\Delta u > 0$ irrespective of α ; but this implies that the function $u = f(x, y)$ has a *minimum* at the point (a, b) . We can similarly show that the function $f(x, y)$ has a *maximum* at the point (a, b) for $\varphi(\alpha) < 0$ ($0 \leq \alpha \leq 2\pi$). Finally, if $\varphi(\alpha)$ assumes both positive and negative values in the interval $(0, 2\pi)$, then let $\varphi(\alpha_1) > 0$ and $\varphi(\alpha_2) < 0$. Assuming that ρ tends to zero while α remains constant, we shall evidently have $\Delta u > 0$ for a small ρ when $\alpha = \alpha_1$ and $\Delta u < 0$ when $\alpha = \alpha_2$. This shows that the function $f(x, y)$ has neither a maximum nor a minimum at the point (a, b) . Hence the sign of the quantity $\varphi(\alpha)$ for $0 \leq \alpha \leq 2\pi$ is, in fact, decisive in determining the character of the given stationary point.

Hence the sign of such a “square trinomial”, *i.e.* the sign of the “discriminant” $\Delta = AC - B^2$, is of decisive importance. We must therefore consider three cases.

1. $\Delta = AC - B^2 > 0$. We have identically

$$A\varphi(\alpha) = (A \cos \alpha + B \sin \alpha)^2 + \Delta \sin^2 \alpha. \quad (1)$$

Since in this case $A \neq 0$, the first term on the right-hand side vanishes only for $\cot \alpha = -B/A$, and the second only for $\sin \alpha = 0$; since these two conditions are incompatible, the angle α has no cotangent for $\sin \alpha = 0$, therefore $A\varphi(\alpha) > 0$ for every α . If $A > 0$,

then $\varphi(\alpha) > 0$ and the function u has a local minimum at the point P ; in contrast if $A < 0$, we have $\varphi(\alpha) < 0$ and P is a local maximum of the function u . Hence if $\Delta > 0$, the point P always gives a local extremum whose nature is determined by the sign of A .

2. $\Delta = AC - B^2 < 0$. Let us at first assume that here also $A \neq 0$. The relation (1) shows that (the first term on the right-hand side is positive and the second equal to zero) we have $A\varphi(\alpha) > 0$ for $\alpha = 0$; on the other hand if

$$\cot \alpha = -\frac{B}{A},$$

we have $A\varphi(\alpha) < 0$ (the first term is equal to zero and the second term negative); hence $\varphi(\alpha)$ has different signs for different values of α and the function u cannot have a local extremum at the point P .

What result do we obtain for $A = 0$? In this case

$$\varphi(\alpha) = 2B \cos \alpha \sin \alpha + C \sin^2 \alpha = \sin \alpha (2B \cos \alpha + C \sin \alpha), \quad (2)$$

where $B \neq 0$, for otherwise we should have $\Delta = 0$. If α is a sufficiently small positive angle, then evidently

$$|C| \sin \alpha < 2|B| \cos \alpha,$$

so that the sign of the bracket $(2B \cos \alpha + C \sin \alpha)$ coincides with the sign of the first term which changes as α is replaced by $-\alpha$; and since $\sin \alpha$ changes its sign during that change, the relation (2) shows that $\varphi(\alpha)$ and $\varphi(-\alpha)$ have opposite signs and the function u cannot have a local extremum at the point P . Hence if $\Delta < 0$, there is no local extremum at the point P .

3. $\Delta = AC - B^2 = 0$. In this case the analysis of terms of the second order in Taylor's formula does not give final results. If the function u has partial derivatives of the third order at the point P , other terms must be analysed in Taylor's formula. However, we shall not consider these problems here.

Example. The function

$$z = x^2 - xy + y^2 - 2x + y$$

has a single stationary point $x = 1, y = 0$ as can be seen by solving the system of equations

$$\frac{\partial z}{\partial x} = 2x - y - 2 = 0, \quad \frac{\partial z}{\partial y} = -x + 2y + 1 = 0.$$

In this case

$$A = 2, \quad B = -1, \quad C = 2,$$

and therefore $\Delta = AC - B^2 = 3$. Since $A > 0$, therefore z has a single extremum, *i.e.* a minimum at the point $(1, 0)$.

For further exercises to § 96 *cf.* Problem Book by B.P. Demidovich, Section VI, Nos. 425, 429, 430, 435, 437, 438.

CHAPTER XXIII

SOME SIMPLE GEOMETRICAL APPLICATIONS OF DIFFERENTIAL CALCULUS

§ 97. Equations of tangent and normal to a plane curve

The geometrical illustration of a derivative as the angular coefficient of the tangent to a given curve at a given point enables us to use the methods of differential calculus in order to solve numerous geometrical problems. Let us assume that we want to draw the tangent to a curve which is the graph of the differentiable function $y = f(x)$ at a point with abscissa a (fig. 55). We know from analytical geometry that the equation of a line which passes through a point with coordinates (a, b) can be written in the form :

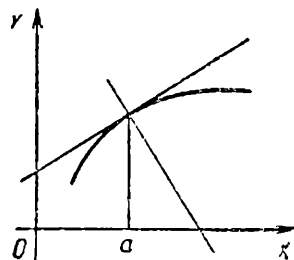


Fig. 55.

$$y - b = k (x - a),$$

where k is the angular coefficient of the straight line. In our case $b = f(a)$, $k = f'(a)$; hence the equation of the tangent to the curve $y = f(x)$ at the point with abscissa a has the form :

$$y - f(a) = f'(a) (x - a).$$

A straight line perpendicular to the tangent at the point of contact is known as *normal* to the given curve at the given point; since the angular coefficients k and k' of two mutually perpendicular lines are connected by the relation $kk' = -1$, the angular coefficient of normal to the curve $y = f(x)$ at the point with abscissa a is equal to $-1 / f'(a)$ (provided $f'(a) \neq 0$). Hence the equation to this normal can be written in the form

$$y - f(a) = -\frac{1}{f'(a)} (x - a),$$

or

$$x - a + f'(a) [y - f(a)] = 0.$$

It is well-known from analytical geometry that it is often more convenient to express the curve in "parametric" form, *i.e.* by means of the following two equations:

$$x = \varphi(t), \quad y = \psi(t),$$

where each value of the "parameter" t in an interval corresponds to a definite point (x, y) on the given curve. We know that the angular coefficient of the tangent to the curve at this point is equal to

$$y' = \frac{dy}{dx}$$

we also know (§ 33) that this expression of the derivative in terms of differentials remains valid when x (and therefore also y) become functions of a new arbitrary variable t as they do in this case. Taking t for the new independent variable we have:

$$y' = \frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}};$$

but $dx / dt = \varphi'(t)$, $dy / dt = \psi'(t)$ and therefore

$$y' = \frac{\psi'(t)}{\varphi'(t)};$$

if we want to write the equation of the tangent to the given curve at a point which corresponds to a value t of the parameter, we must take into consideration the fact that the coordinates of this point are equal to $x = \varphi(t)$, $y = \psi(t)$ and the angular coefficient of the tangent is equal to $y' = \psi'(t) / \varphi'(t)$; hence the equation of the tangent is as follows:

$$y - \psi(t) = \frac{\psi'(t)}{\varphi'(t)} [x - \varphi(t)],$$

or

$$\frac{x - \varphi(t)}{\varphi'(t)} = \frac{y - \psi(t)}{\psi'(t)}, \quad (1)$$

which owing to its symmetry is very convenient; since the angular coefficient of the normal is in this case equal to

$$-\frac{1}{y'} = -\frac{\varphi'(t)}{\psi'(t)},$$

therefore the equation of the normal is as follows :

$$y - \psi(t) = -\frac{\varphi'(t)}{\psi'(t)} [x - \varphi(t)],$$

or

$$\varphi'(t) [x - \varphi(t)] + \psi'(t) [y - \psi(t)] = 0. \quad (2)$$

If the curve is given in terms of polar coordinates

$$r = f(\theta)$$

and we wish to write the equation of the tangent at the point with coordinates $\theta_0, r_0 = f(\theta_0)$, we can solve this problem in the same way as the previous one remembering that cartesian and polar coordinates are connected by the relation.

$$x = r \cos \theta, \quad y = r \sin \theta$$

and for points on the given curve it takes the form

$$x = f(\theta) \cos \theta, \quad y = f(\theta) \sin \theta. \quad (3)$$

Equation (3) represents the given curve in parametric form, where the polar angle θ is a parameter ; we have :

$$\frac{dx}{d\theta} = f'(\theta) \cos \theta - f(\theta) \sin \theta, \quad \frac{dy}{d\theta} = f'(\theta) \sin \theta + f(\theta) \cos \theta.$$

According to (1) the equation of the tangent at the point $\theta = \theta_0$ can be written in cartesian coordinates in the following form:

$$\frac{y - f(\theta_0) \sin \theta_0}{f'(\theta_0) \sin \theta_0 + f(\theta_0) \cos \theta_0} = \frac{x - f(\theta_0) \cos \theta_0}{f'(\theta_0) \cos \theta_0 - f(\theta_0) \sin \theta_0},$$

and in polar coordinates in the form

$$\frac{r \sin \theta - f(\theta_0) \sin \theta_0}{f'(\theta_0) \sin \theta_0 + f(\theta_0) \cos \theta_0} = \frac{r \cos \theta - f(\theta_0) \cos \theta_0}{f'(\theta_0) \cos \theta_0 - f(\theta_0) \sin \theta_0},$$

or, assuming that $f(\theta_0) = r_0, f'(\theta_0) = r'_0$, in the form

$$\frac{r \sin \theta - r_0 \sin \theta_0}{r'_0 \sin \theta_0 + r_0 \cos \theta_0} = \frac{r \cos \theta - r_0 \cos \theta_0}{r'_0 \cos \theta_0 - r_0 \sin \theta_0}.$$

For exercises to § 97 cf. Problem Book by B.P. Demidovich, Section II, Nos. 119—121, 124, 126, 141, 142.

§ 98. Tangential line and normal plane to a curve in space

The geometrical definition of a tangent to a curve in space does not differ from its definition relating to plane curves. If we wish to draw a tangent to the given curve at a given point M , we take another point N on the curve close to M and draw a straight line (chord) through these two points; if, when the point N comes indefinitely close to the point M , the drawn chord tends to a definite limiting position, then this limiting straight line is a tangent to the given curve at the point M . Assuming that the curve in space is given by parametric equations of the form

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t),$$

we must try to find the equation of the tangent to this curve at a point corresponding to the given value of t_0 of the parameter t (i.e. at a point with coordinates $x_0 = \varphi(t_0)$, $y_0 = \psi(t_0)$, $z_0 = \chi(t_0)$). Let the parameter t receive an increment Δt and let us go from the given point to the point

$$x_0 + \Delta x = \varphi(t_0 + \Delta t), \quad y_0 + \Delta y = \psi(t_0 + \Delta t), \quad z_0 + \Delta z = \chi(t_0 + \Delta t).$$

The straight line (chord) which connects the two points (the given and the displaced point) can, in accordance with the laws of analytical geometry, be expressed by the following equations:

$$\frac{x - x_0}{\Delta x} = \frac{y - y_0}{\Delta y} = \frac{z - z_0}{\Delta z},$$

or by an equivalent system

$$\frac{x - x_0}{\frac{\Delta x}{\Delta t}} = \frac{y - y_0}{\frac{\Delta y}{\Delta t}} = \frac{z - z_0}{\frac{\Delta z}{\Delta t}}. \quad (1)$$

If we assume that Δt tends to zero and the functions $\varphi(t)$, $\psi(t)$ and $\chi(t)$ have nonzero derivatives at $t = t_0$, which we shall respectively denote by x'_0 , y'_0 , z'_0 , then the system of equations (1) for the drawn chord has the following form in the limit:

$$\frac{x - x_0}{x'_0} = \frac{y - y_0}{y'_0} = \frac{z - z_0}{z'_0}, \quad (2)$$

or, which is the same,

$$\frac{x - \varphi(t_0)}{\varphi'(t_0)} = \frac{y - \psi(t_0)}{\psi'(t_0)} = \frac{z - \chi(t_0)}{\chi'(t_0)}. \quad (3)$$

The system of equations (2) or (3) is analogous to equation (1) § 97 and evidently provides the analytical expression for the tangent to a curve in space.

Denoting by α , β and γ the angles made by the tangent to the given curve at the point (x_0, y_0, z_0) with the positive direction of the axes of coordinates, we have in accordance with the laws of analytical geometry :

$$\begin{aligned}\cos \alpha &= \frac{\varphi'(t_0)}{\sqrt{\varphi'^2(t_0) + \psi'^2(t_0) + \chi'^2(t_0)}}, \quad \cos \beta = \frac{\psi'(t_0)}{\sqrt{\varphi'^2(t_0) + \psi'^2(t_0) + \chi'^2(t_0)}}, \\ \cos \gamma &= \frac{\chi'(t_0)}{\sqrt{\varphi'^2(t_0) + \psi'^2(t_0) + \chi'^2(t_0)}}.\end{aligned}\quad (t_0)$$

In particular, if the given curve is expressed by the following equations :

$$y = y(x), \quad z = z(x), \quad (4)$$

the equations of the tangent at the point (x_0, y_0, z_0) have the form :

$$x - x_0 = \frac{y - y_0}{y'(x_0)} = \frac{z - z_0}{z'(x_0)},$$

and we have :

$$\begin{aligned}\cos \alpha &= \frac{1}{\sqrt{1 + y'^2(x_0) + z'^2(x_0)}}, \quad \cos \beta = \frac{y'(x_0)}{\sqrt{1 + y'^2(x_0) + z'^2(x_0)}}, \\ \cos \gamma &= \frac{z'(x_0)}{\sqrt{1 + y'^2(x_0) + z'^2(x_0)}}.\end{aligned}$$

In all cases the choice of the sign in front of the radical depends on the choice of one or other direction of the tangent.

A plane drawn through a point of a curve in space perpendicularly to the tangent at that point is known as the *normal plane* to the given curve at the given point. Normal planes are very important in the theory of curves in space and play a similar role as normal lines (*i.e.* ordinary normals) in relation to plane curves. In accordance with the general laws of analytical geometry we can, by knowing the equations of the tangent in the forms (2) or (3), write directly the equation of a normal plane to the same curve at the same point in the following form :

$$x'_0(x - x_0) + y'_0(y - y_0) + z'_0(z - z_0) = 0,$$

or

$$\varphi'(t_0)[x - \varphi(t_0)] + \psi'(t_0)[y - \psi(t_0)] + \chi'(t_0)[z - \chi(t_0)] = 0.$$

We can see that these equations are analogous to the equation (2) § 97 of a normal to a plane curve.

If the curve is expressed by equations of the form (4), the equation of the normal plane at the point (x_0, y_0, z_0) has the form:

$$x - x_0 + y'(x_0)(y - y_0) + z'(x_0)(z - z_0) = 0.$$

For exercises to § 98 cf. Problem Book by B. P. Demidovich, Section VI, Nos. 341, 342, 344, 346.

§ 99. Tangential and normal planes to a surface

Let us consider a surface in space which is expressed by the following equation:

$$F(x, y, z) = 0 \quad (1)$$

and choose an arbitrary point M with coordinates x_0, y_0, z_0 , on it so that $F(x_0, y_0, z_0) = 0$. Let us draw an arbitrary curve on the surface (1) which passes through the point M ; let this curve have the following parametric equations:

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t). \quad (2)$$

Since the curve (2) lies wholly on the surface (1), we must have identically (*i.e.* for every arbitrary value of the parameter t in some region)

$$F[\varphi(t), \psi(t), \chi(t)] \equiv 0. \quad (3)$$

On the other hand, since the curve passes through the point $M(x_0, y_0, z_0)$, therefore for a given value t_0 of the parameter t we have:

$$x_0 = \varphi(t_0), \quad y_0 = \psi(t_0), \quad z_0 = \chi(t_0).$$

In order to draw further conclusions we must now assume that the function $F(x, y, z)$ is differentiable at the point $M(x_0, y_0, z_0)$. In chapter 22 we have agreed to call the function $u = f(x, y, z)$ differentiable at the point (x, y, z) if, assuming that

$$\Delta u = f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z),$$

$$du = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y + \frac{\partial f}{\partial z} \Delta z,$$

$$\rho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2},$$

we have for $\rho \rightarrow 0$:

$$\Delta u = du + o(\rho).$$

The law for differentiating composite functions holds for differentiable functions (§ 92) : if the function $u = f(x, y, z)$ is differentiable and x, y, z , are also differentiable functions of a new variable t , then

$$\frac{du}{dt} = \frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt} + \frac{\partial u}{\partial z} \frac{dz}{dt}. \quad (4)$$

We have already agreed to assume in our case that the function $F(x, y, z)$ is differentiable at the point (x_0, y_0, z_0) which corresponds to the value t_0 of the parameter t . If we express x, y, z in terms of t by means of the relation (2), then $F(x, y, z)$ becomes a function of the parameter t and, according to (3), it is a constant. Hence $dF/dt = 0$ and it follows from formula (4) that

$$\frac{\partial F}{\partial x} \frac{dx}{dt} + \frac{\partial F}{\partial y} \frac{dy}{dt} + \frac{\partial F}{\partial z} \frac{dz}{dt} \equiv 0. \quad (5)$$

Having made this remark we can now write the equation of the tangent to the curve (2) at the point $M(x_0, y_0, z_0)$ to correspond to the value t_0 of the parameter t . According to formula (2) § 98 these equations can be written in the form

$$\frac{x - x_0}{x'_0} = \frac{y - y_0}{y'_0} = \frac{z - z_0}{z'_0}.$$

Hence for every point (x, y, z) on our tangent these three relations have the same value which we shall conveniently denote by $1/\lambda$ (where λ is clearly different for different points on the tangent). But in that case

$$x'_0 = \lambda(x - x_0), \quad y'_0 = \lambda(y - y_0), \quad z'_0 = \lambda(z - z_0). \quad (6)$$

On the other hand, the identity (5) gives us for $t = t_0$, if we denote respectively by A, B and C the values $\partial F / \partial x, \partial F / \partial y$ and $\partial F / \partial z$ at the point $M(x_0, y_0, z_0)$:

$$Ax'_0 + By'_0 + Cz'_0 = 0.$$

Substituting in this relation x'_0, y'_0 and z'_0 by their expressions from (6) and cancelling λ we obtain :

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \quad (7)$$

Let us note that in this case x, y, z are coordinates of any point on the tangent to the curve (2) at the point M and A, B and C are values of the partial derivatives of the function F at the point M . If we regard x, y, z in the equation (7) as current coordinates, then the latter equation represents a plane which passes through the point M ; it depends on the form of the surface (1) but is quite independent of the curve (2) drawn on this surface. The fact that the coordinates of every point on the tangent to the curve (2) are connected by the equation (7) show, that this tangent lies wholly in the plane (7). But the curve (2) is an arbitrary curve drawn on the surface (1) through the point M . We can draw an infinite number of such curves; we thus see that the tangents to all those curves lie in the same plane (7); this plane which is a sort of "carrier" (a geometrical set of points) of tangents to arbitrary curves drawn on the surface (1) through the point M is known as the *tangential plane* to the given surface at the point M . The equation (7) of tangential plane can be written in a more expressive form if instead of the neutral notation A, B, C for the partial derivatives of the function F at the point M we write respectively

$$\frac{\partial F}{\partial x_0}, \quad \frac{\partial F}{\partial y_0}, \quad \frac{\partial F}{\partial z_0},$$

where the index 0 shows that all three derivatives are taken at the point $x = x_0, y = y_0, z = z_0$. The equation of the tangential plane thus becomes:

$$\frac{\partial F}{\partial x_0} (x - x_0) + \frac{\partial F}{\partial y_0} (y - y_0) + \frac{\partial F}{\partial z_0} (z - z_0) = 0.$$

The straight line drawn through the point M perpendicular to the tangential plane is said to be *normal* to the surface at the point M . According to the laws of analytical geometry the equations of this straight line (when none of the three partial derivatives vanishes) can be written in the following form:

$$\frac{x - x_0}{\frac{\partial F}{\partial x_0}} = \frac{y - y_0}{\frac{\partial F}{\partial y_0}} = \frac{z - z_0}{\frac{\partial F}{\partial z_0}}.$$

In particular, if the surface is expressed by the equation:

$$z = f(x, y), \tag{8}$$

we have $F(x, y, z) = z - f(x, y)$ and the equation of the tangential plane becomes

$$z - z_0 = \frac{\partial f}{\partial x_0} (x - x_0) + \frac{\partial f}{\partial y_0} (y - y_0),$$

and the equation of the normal becomes

$$z - z_0 = - \frac{x - x_0}{\frac{\partial f}{\partial x_0}} = - \frac{y - y_0}{\frac{\partial f}{\partial y_0}}.$$

If we denote respectively by α , β and γ the angles between the normal to the surface (8) at the point (x_0, y_0, z_0) and the positive direction of the axes of coordinates, then, as we know from analytical geometry, we have

$$\cos \alpha = \frac{-\frac{\partial f}{\partial x_0}}{\pm \sqrt{1 + \left(\frac{\partial f}{\partial x_0}\right)^2 + \left(\frac{\partial f}{\partial y_0}\right)^2}},$$

$$\cos \beta = \frac{-\frac{\partial f}{\partial y_0}}{\pm \sqrt{1 + \left(\frac{\partial f}{\partial x_0}\right)^2 + \left(\frac{\partial f}{\partial y_0}\right)^2}},$$

$$\cos \gamma = \frac{1}{\pm \sqrt{1 + \left(\frac{\partial f}{\partial x_0}\right)^2 + \left(\frac{\partial f}{\partial y_0}\right)^2}}.$$

The choice of the sign in front of the radical depends on our choice of direction for the normal under consideration; it is self-evident that this sign must be the same in all three formulae.

For exercises to § 99 cf. Problem Book by B.P. Demidovich, Section VI, Nos. 351, 352, 360-362.

§ 100. Direction of convexity and concavity of a curve

We shall now return to the theory of plane curves and deal with another problem, *viz.* the direction of convexity and concavity of a curve. Let us assume that the function $y = f(x)$ has derivatives of first two orders at $x = a$. The equation of the tangent at the point a to the curve $y = f(x)$ has the form

$$y = f(a) + f'(a)(x - a).$$

At the point $a + h$, close to a , the ordinate of the tangent will therefore be

$$y_{tan} = f(a) + hf'(a),$$

whereas the ordinate of the curve at that point is equal to

$$y_{curve} = f(a + h);$$

in order to find out which of these lines lies above the other in the immediate neighbourhood of the point a we construct the difference

$$y_{curve} - y_{tan} = f(a + h) - f(a) - hf'(a),$$

and since, according to our assumption, $f''(a)$ exists, it follows from Taylor's formula that we can represent the right-hand side of this equation in the form

$$\frac{h^2}{2} f''(a) + o(h^2),$$

and we obtain:

$$y_{curve} - y_{tan} = \frac{h^2}{2} f''(a) + o(h^2).$$

Let us assume that $f''(a) \neq 0$; in that case the second term on the right-hand side is infinitely small as compared to the first term for $h \rightarrow 0$; the sign of the right-hand side (and therefore also of the left-hand side) coincides with the sign of the first term when $|h|$ is infinitely small, *i.e.* with the sign of $f''(a)$. If $f''(a) > 0$, then $y_{curve} > y_{tan}$ at all points sufficiently close to a , *i.e.* the curve lies above its tangent (fig. 56, *a*); the position of the two is reversed for $f''(a) < 0$ (fig. 56, *b*).

In the first case it is said that the curve $y = f(x)$ has a convexity from below at the point a (in the direction of negative y) or a concavity from above (in the direction of positive y); in the second case the position is reversed — the curve is convex from above and concave from below at the point a . The use of this terminology is at once apparent by looking at figs. 56 *a* and *b*. We thus see that the sign of the second derivative determines the direction of convexity and concavity of a curve in the same way as the sign of the first derivative tells us whether the function increases or decreases.

Conversely, we know that the curve $y = f(x)$ is convex from below in the neighbourhood of the point a (i.e. it lies above its tangent). If $f''(a)$ exists, then it follows from above that it cannot be negative, for in that case the relative positions of the tangent and curve would be reversed. Therefore $f''(a) \geq 0$. The case when $f''(a) = 0$ is quite possible as can be seen on the example of the curve $y = x^4$ at $x = 0$; similar arguments clearly show that when curve is convex from above at the point a , we should have $f''(a) \leq 0$ and here again the value $f''(a) = 0$ is possible. Finally, in the neighbourhood of the point a the curve may lie above its tangent to one side of a and below it to the other side of a ; this is the position with the function $y = x^3$ in the neighbourhood of the point $x = 0$, which we have already considered on several occasions (fig. 21 § 40); the curve intersects its tangent at such a point and changes the direction of its convexity in the process.

Points of this type are usually known as *inflexions* of the given curve. It is evident that the derivative of second order $f''(a)$, in case it exists, should be equal to zero at a point of inflexion.

Hence if $f''(a) = 0$, we cannot draw any conclusions as to the direction of convexity of the curve in the neighbourhood of the point a ; we may in this case have a convexity from above; a convexity from below or a point of inflexion; even more complicated cases are possible. For further analysis it is necessary to investigate successive terms in Taylor's series in the same way as we have done while finding extrema of functions (§ 41). However, we shall not go into further details here.

For exercises to § 100 cf. Problem Book by B. P. Demidovich, Section II, Nos. 348, 349, 352-354, 362.

§ 101. Curvature of a plane curve

It is obvious that different curves can have different curvatures in different intervals. The curve shown in fig. 57 has an almost recti-

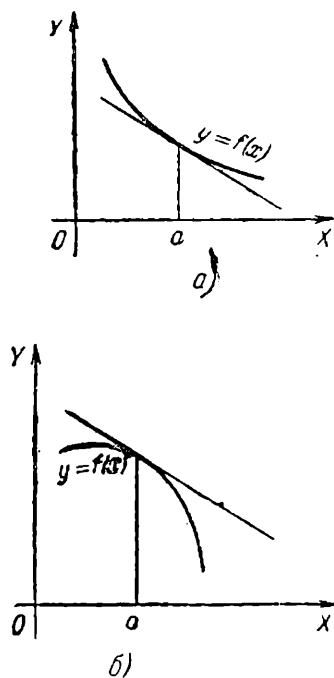


Fig. 56.

linear left side and there its curvature is naught. A circle has the same degree of curvature throughout. If we imagine several circles of different radii having a common tangent at a point (fig. 58), then we can see clearly that the degree of curvature is greater, the smaller the radius of the circle is. A large curvature of a path (bend) can be experienced without seeing it (without looking out of the window) while travelling by car or rail. The importance of this type of curvature prompts us to evaluate it scientifically; we must learn to compare not only the qualitative curvatures of various curves (one curve has, say, a greater curvature than the other) but we must learn

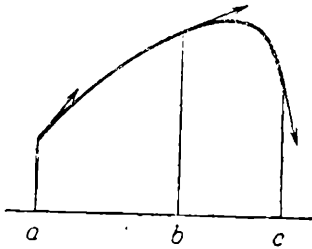


Fig. 57.

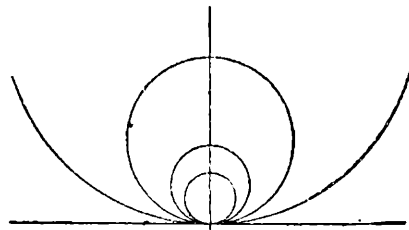


Fig. 58.

to give it a quantitative evaluation, *i.e.* we must learn to *measure* the degree of curvature of curves. In the science of road construction (any kind of road or rail) this exact approach is essential and this problem is also important in physics and thermodynamics; we cannot merely say that one body is hotter or cooler than another body, but we must learn to express quantitatively the heat of various bodies (*i.e.* their temperature).

In our visual representation the degree of curvature of a curve is closely related to the rate at which it changes its direction. Thus the curve shown in fig. 57 hardly changes its direction in the interval (a, b) — the tangents at different points (and in particular at points with abscissae a and b) are almost parallel to one another; therefore this curve appears to have a small curvature in the interval (a, b) ; in contrast, this curve appears to have a large curvature in the interval (b, c) , for its direction changes considerably in that interval; thus the tangents drawn at the ends of this interval have very different directions. It is therefore clear that in order to measure the degree of curvature of a curve in an interval quantitatively we must take for the basic value the angle by which the tangent to the given curve rotates in passing through this interval, *i.e.* the angle between the tangents at the beginning and end of the given interval. However, if only one angle is given, we cannot determine the degree of curvature of the curve in the given interval. In fact, if you are told that,

say, the railway line has a bend of 30° at a certain spot, this will tell nothing about the curvature of this path. You will undoubtedly ask how long this section is in which this bend takes place. If, for example, the path has turned by 30° over a distance of 2 km, then the curvature is very insignificant; if, however, the path has bent to the same extent over a distance of 100 m, then this curvature is large. Thus in order to evaluate the degree of curvature of the path in a given interval it is necessary to know not only the angle φ which measures the change in direction of the curve in the given interval but also the length s of this interval. In this case the natural measure of the degree of curvature of the curve in the given interval will evidently be given by the ratio φ / s , i.e. by *the change in direction per unit length of path*. This quantity is known as *average curvature* of the curve in the given interval. The meaning of this term is self-evident: it is obvious that a curve can have different curvatures in different parts of a given interval and the concept of average curvature does not take all these differences into account; it merely shows the average change in direction of the curve per unit length of its path.

If now we want to consider not the average but *local* characteristic of the degree of curvature in the immediate neighbourhood of a certain point A on the curve, we must again use the same arguments as in § 26 when we were considering the instantaneous (local) velocity of motion at a given instant on the basis of the average velocity of a body in a given time interval. Let us take another point B on our curve apart from the point A ; let the length of the arc AB of the given curve be equal to s and the angle between the tangents to the curve at the points A and B be equal to φ so that the average curvature of the arc AB is equal to φ/s . If the point B lies close to A (i.e. if s is small), we have reason to believe that the degree of curvature of the curve will not change too much in transition from A to B and that therefore the average curvature φ / s of the arc AB will still give us a sufficiently accurate description of the degree of curvature of the curve in the immediate neighbourhood of the point A ; this characteristic will be the more accurate the smaller s is, i.e. the smaller the distance from the point B to the point A is. Therefore if, as $B \rightarrow A$ (or, which is the same, as $s \rightarrow 0$) the average curvature φ / s tends to a limit K , then we can naturally regard this limit as the (local) *curvature of our curve at the point A* (or in the immediate neighbourhood of the point A).

Hence *curvature of the given curve at the given point A* is expressed by the limit of the ratio of the angle between the tangents of the curve at the points

A and B to the length of the arc AB provided the point B lies on the given curve and approaches indefinitely close to the point A.

Having established the geometrical meaning of the concept of (local) curvature we will now show how the methods of differential calculus enable us to evaluate the curvature of a given curve at any point. Let the function $y = f(x)$ which is represented graphically by the given curve have derivatives of the first two orders at the point x . Let us consider a point $B(x + \Delta x, y + \Delta y)$ on the given

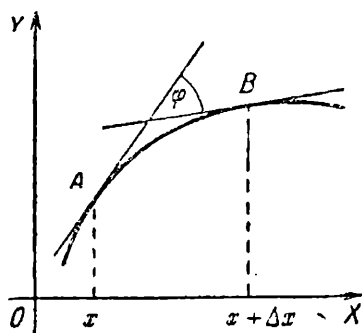


Fig. 59.

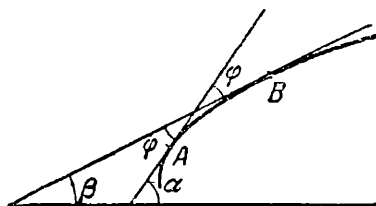


Fig. 60.

curve other than the point $A(x, y)$ (fig. 59). If we denote the angles made by the tangents to the curve at these two points and the positive direction of the OX -axis by α and β respectively, then evidently

$$\tan \alpha = f'(x), \quad \tan \beta = f'(x + \Delta x).$$

The angle φ between these two tangents is equal to $\varphi = |\alpha - \beta| = |\arctan f'(x + \Delta x) - \arctan f'(x)|$, as can be seen from fig. 60. On the other hand, if we denote the length of the arc of the given curve in the interval (a, x) by $s(x)$, where a is a constant, then the length s of the section AB of the given curve is equal to

$$s = s(x + \Delta x) - s(x).$$

We thus obtain the following expression for the average curvature of the arc AB of the given curve :

$$\frac{\varphi}{s} = \frac{|\arctan f'(x + \Delta x) - \arctan f'(x)|}{s(x + \Delta x) - s(x)}$$

or, dividing the numerator and the denominator by Δx

$$\frac{\varphi}{s} = \frac{\frac{|\arctan f'(x + \Delta x) - \arctan f'(x)|}{\Delta x}}{\frac{s(x + \Delta x) - s(x)}{\Delta x}}.$$

If we now assume that Δx tends to zero, then it follows from our assumption of existence of $y'' = f''(x)$ that the numerator will tend to

$$\left| \frac{d \arctan y'}{dx} \right| = \frac{|y''|}{1 + y'^2},$$

whereas, according to § 52, the denominator will have a positive limit

$$\frac{ds}{dx} = \sqrt{1 + y'^2}.$$

We thus obtain the following expression for the curvature of the given curve at the point $A(x, y)$:

$$K = \lim_{s \rightarrow 0} \frac{\varphi}{s} = \frac{|y''|}{(1 + y'^2)^{\frac{3}{2}}}, \quad (1)$$

and our problem is solved.

If the curve is given by a parametric equation

$$x = \varphi(t), \quad y = \psi(t),$$

then

$$y' = \frac{dy}{dx} = \frac{\psi'(t)}{\varphi'(t)},$$

$$y'' = \frac{d}{dx} \left[\frac{\psi'(t)}{\varphi'(t)} \right] = \frac{\frac{d}{dt} \left[\frac{\psi'(t)}{\varphi'(t)} \right]}{\frac{dx}{dt}} = \frac{\varphi'(t) \psi''(t) - \varphi''(t) \psi'(t)}{[\varphi'(t)]^3},$$

and after elementary calculations we obtain from formula (1):

$$K = \frac{|\varphi'(t) \psi''(t) - \psi'(t) \varphi''(t)|}{[\varphi'^2(t) + \psi'^2(t)]^{\frac{3}{2}}}. \quad (2)$$

A straight line is expressed by a linear equation $y = mx + n$; therefore $y'' = 0$ at every point and it follows from (1) that $K = 0$; hence the curvature of a straight line is equal to zero. In the case of a circle of radius r it is more convenient to use the parametric formula

$$x = r \cos t, \quad y = r \sin t;$$

after elementary calculations we obtain from formula (2) $K = 1/r$; hence the curvature of a circle is the same at every point and is equal to the reciprocal value of its radius.

§ 102. Tangential circle

Let the curve $y = f(x)$ have a nonzero curvature K at the point $A(x, y)$ ($y'' \neq 0$). Let us draw a normal to the curve at the point A (fig. 61) and mark on it an interval AC of length $1/K$ in the direction of the concavity of the curve. If we now draw a circle with centre at C and radius $r = 1/K$, then this circle will pass through the point A and have a common tangent with the curve at that point; since the radius CA lies on the normal to the curve; moreover, the convexities of the curve $y = f(x)$ and of our circle are in the same direction at that point; also the curvature of these curves at the point A is the same, since the curvature of our circle is equal to $1/r = K$. We can therefore say that among all circles which can be drawn through the point A our circle comes closest of all to the direction of the circle at the point A having a common direction with it (tangent) and the same curvature; its convexity is also directed to the same.

This circle is known as the *circle of curvature* of the given curve at the point A . Its radius

$$r = \frac{1}{K} = \frac{(1+y'^2)^{\frac{3}{2}}}{|y''|}$$

is known as the *radius of curvature* and its centre as the *centre of curvature* of the curve $y = f(x)$ at the point A . In the same way as a

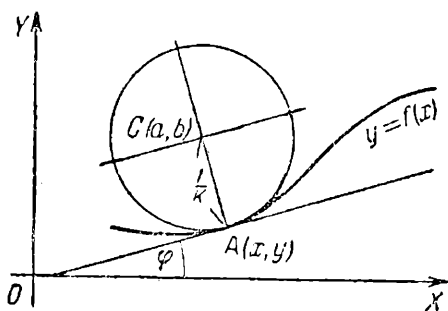


Fig. 61.

tangent can be used instead of a curve in all instances where only the direction of the curve is involved, so the circle of curvature can replace the given curve in all problems where apart from the direction of the curve its curvature and direction of convexity are also involved. This is the main part played by the circle of curvature in many geometrical investigations which involve

the given curve. It thus becomes obvious why the circle of curvature is called a *tangential* (or, as it is often said, a *contiguous*) circle to the given curve at the point A .

Let us now find an analytical expression for the coordinates (a, b) of the centre of curvature. In the case shown in fig. 61 $|y''| = y'' > 0, y' > 0, x > a, y < b$. The differences $x - a$ and $b - y$ are the projections of the interval r in the directions OX and OY res-

pectively. Therefore if we denote the angle between the tangent at the point A and the positive direction of the OX -axis by φ (so that $\tan \varphi = y'$), then

$$x - a = r \sin \varphi = \frac{(1 + y'^2)^{\frac{3}{2}}}{y''} \cdot \frac{y'}{\sqrt{1 + y'^2}} = \frac{y' (1 + y'^2)}{y''},$$

$$b - y = r \cos \varphi = \frac{(1 + y'^2)^{\frac{3}{2}}}{y''} \cdot \frac{1}{\sqrt{1 + y'^2}} = \frac{1 + y'^2}{y''},$$

hence

$$a = x - \frac{y' (1 + y'^2)}{y''}, \quad b = y + \frac{1 + y'^2}{y''}.$$

It can be readily shown that these formulae remain valid if the course of the curve $y = f(x)$ in the neighbourhood of the point A is different.

The concept of a tangential circle can also be approached from another point of view which shows even more clearly its analogy with a tangent. We have defined a tangent as the limiting position of the chord joining the given point $A(x, y)$ and another point $B(x + \Delta x, y + \Delta y)$ on the same curve when Δx (and therefore also Δy) tends to zero. If, however, we want to draw through these two points not a straight line but a circle, we encounter difficulties since an infinite number of such circles can be drawn: we know that in order to define a circle uniquely we must be given not two but three points (not lying on the same straight line) through which the circle is to be drawn. Therefore, apart from the point A we shall take two more points B_1 and B_2 on our curve with abscissae x_1 and x_2 . The equation of the circle drawn through the points A , B_1 and B_2 can be written in the form

$$(x - \alpha)^2 + (y - \beta)^2 = \rho^2,$$

where the radius ρ and the coordinates α and β of the centre of the circle are determined from the condition that the circle given by this equation must pass through the points A , B_1 and B_2 . If we assume that

$$(x - \alpha)^2 + [f(x) - \beta]^2 - \rho^2 = F(x),$$

then this condition can evidently be written in the form

$$F(x) = 0, \quad F(x_1) = 0, \quad F(x_2) = 0. \quad (1)$$

These three equations enable us to determine the unknowns α , β and ρ . However, we shall use another method. Let us assume that $x < x_1 < x_2$. It then follows from (1) that we can apply Rolle's theorem to each of the intervals (x, x_1) and (x_1, x_2) . This gives us :

$$F'(\xi_1) = F'(\xi_2) = 0;$$

where $x < \xi_1 < x_1 < \xi_2 < x_2$; therefore, applying Rolle's theorem to the function $F'(x)$ in the interval (ξ_1, ξ_2) we obtain :

$$F''(\xi) = 0,$$

where ξ lies between ξ_1 and ξ_2 .

Let us now assume that the points B_1 and B_2 approach indefinitely close to the point A along the given curve, *i.e.* we assume that $x_1 \rightarrow x$ and $x_2 \rightarrow x$; it is then evident that the points ξ_1 , ξ_2 and ξ will tend to x . Our circle drawn through the points A , B_1 and B_2 will in this process continuously change its radius and its direction; for every position of the points B_1 and B_2 we can find from the equation (1) the elements α , β and ρ for this circle; we could perform these calculations and find the limits to which α , β and ρ tend as $x_1 \rightarrow x$ and $x_2 \rightarrow x$. However, it is simpler to act otherwise. Since we have for the arbitrary x_1 and x_2

$$F(x) = 0, \quad F'(\xi_1) = 0, \quad F''(\xi) = 0$$

and since $\xi \rightarrow x$ as $\xi_1 \rightarrow x$, therefore we should have for a limiting circle *)

$$F(x) = F'(x) = F''(x) = 0,$$

i.e. denoting the elements of the limiting circle by a , b and r and assuming that $f(x) = y$

$$F(x) = (x - a)^2 + (y - b)^2 - r^2 = 0,$$

$$F'(x) = 2(x - a) + 2(y - b)y' = 0,$$

$$F''(x) = 2 + 2y'^2 + 2(y - b)y'' = 0.$$

The last of these equations gives us directly :

$$b - y = \frac{1 + y'^2}{y''},$$

*) In this case we assume continuity of $F''(x)$ at the point x_1 ; for this to be the case that $f''(x)$ should be continuous at that point.

after which we obtain from the last but one equation :

$$a - x = - \frac{(1 + y'^2) y'}{y''};$$

finally after substituting in the first of the above equations the values of $b - y$ and $a - x$ we obtain :

$$r = \frac{(1 + y'^2)^{\frac{3}{2}}}{|y''|}.$$

The formulae obtained above show that the required limiting circle does, in fact, coincide with the circle of curvature of the given curve at the point A .

Thus *the circle of curvature (or the tangential circle) of the given curve at the point A is the limiting position of the circle which passes through the point A and through two other points which lie infinitely close to the given point on the given curve.*

For exercises to §§ 101—102 cf. Problem Book by B.P. Demidovich, Section II, Nos. 566, 567, 571, 572, 575, 576, 577.

CHAPTER XXIV

IMPLICIT FUNCTIONS

§ 103. The simplest problem

We have already met implicit functions in § 92 and we shall now consider the problem solved there. We worked on the assumption that $y = f(x)$ identically satisfies the following equation in an interval (a, b)

$$F(x, y) = 0, \quad (1)$$

i.e. we had

$$F[x, f(x)] = 0 \quad (a \leq x \leq b).$$

Having assumed that the function $f(x)$ is differentiable in the interval (a, b) we tried to express its derivative in terms of the partial derivatives of the function F with respect to x and y and found the following expression for this derivative :

$$y' = f'(x) = - \frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}.$$

However, in concrete cases the problem is usually somewhat different. Only the function $F(x, y)$ is given ; as for as the function $y = f(x)$ which satisfies the equation (1) in an interval is concerned, neither its continuity and differentiability nor even its existence is assumed beforehand ; on the contrary the determination of the conditions for existence of such a function and study of its properties is the main purpose of our problem. Among the numerous methods used for the determination of new non-elementary functions this "implicit" description of functions by equations plays a very important part. The set of laws which characterise this method of

($k = 1, 2, \dots, m$); if these functions f_k exist, we must study their properties (continuity, differentiability, etc.).

In future we shall understand by the *simplest* problem a problem where only one determining equation is given (regardless of the number of variables) and by the *general* problem—a problem where several determining equations are given. In this paragraph we shall only study the simplest problem. We shall see that the method of investigation does not depend on the number of variables; therefore in order to prevent our notation from becoming too complicated we shall consider the simplest case involving two variables, although all arguments used will remain valid for any number of variables.

Thus we shall assume that we are given the equation

$$F(x, y) = 0 \quad (3)$$

and try to find the function $y = f(x)$ which identically satisfies this equation in very region of values of x . It is clear at once that

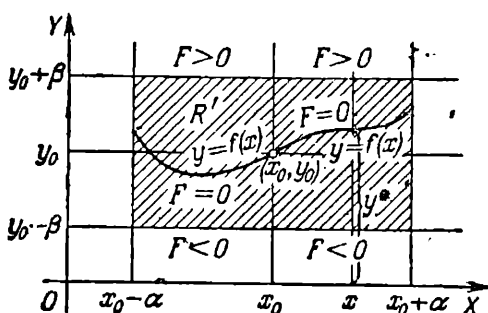


Fig. 62

existence of this function as well as its properties depend on the properties of the given function $F(x, y)$. The more assumptions we make with regard to the function F , the more definite statements we can make with regard to the function $f(x)$. Hence our problem can have several variations; here we shall only consider one variation, i.e.

a variation which crops up most often in applications of the theory of implicit functions.

Theorem 1. *Let the function $F(x, y)$ be continuous and have continuous partial derivatives with respect to both variables in a rectangle*

$$R(x_0 - a \leq x \leq x_0 + a, y_0 - b \leq y \leq y_0 + b).$$

$$\text{Let.} \quad F(x_0, y_0) = 0, \quad F'_y(x_0, y_0) \neq 0.$$

Then there exists a unique function $y = f(x)$ which is continuous and satisfies the equation (3) in some interval $\Delta(x_0 - \alpha \leq x \leq x_0 + \alpha)$; it is equal to y_0 for $x = x_0$. This function has a continuous derivative in the interval Δ .

Proof. 1°. *Definition of the function $f(x)$.*

Let us assume that $F'_y(x_0, y_0) > 0$. According to the lemma § 23 (which holds for all continuous functions irrespective of the number of variables) we should have $F'_y(x, y) > 0$ also for all points of a rectangle $R'(x_0 - \alpha \leq x \leq x_0 + \alpha, y_0 - \beta \leq y \leq y_0 + \beta)$ (fig. 62). It is important to remember that the numbers α and β can be arbitrarily small; hence we are already in a position to assume that R' lies entirely within R . We therefore have:

$$F'_y(x_0, y) > 0 \quad (y_0 - \beta \leq y \leq y_0 + \beta),$$

which shows that $F(x_0, y)$ is an increasing function of y in the interval $(y_0 - \beta, y_0 + \beta)$; and since $F(x_0, y_0) = 0$, therefore

$$F(x_0, y_0 - \beta) < 0, \quad F(x_0, y_0 + \beta) > 0. \quad (4)$$

According to the lemma § 23 these inequalities remain valid when we replace x_0 by an arbitrary number x which is sufficiently close to x_0 ; and since we have agreed above that α can be arbitrarily small, we have every justification to assume that the inequalities

$$F(x, y_0 - \beta) < 0, \quad F(x, y_0 + \beta) > 0 \quad (5)$$

are satisfied at every point x in the interval $\Delta(x_0 - \alpha, x_0 + \alpha)$. Let us choose and temporarily fix an arbitrary point x in this interval and assume that

$$F(x, y) = \varphi(y) \quad (y_0 - \beta \leq y \leq y_0 + \beta);$$

we have:

$$\varphi'(y) = F'_y(x, y) > 0 \quad (y_0 - \beta \leq y \leq y_0 + \beta),$$

and since the function $\varphi(y)$ increases in the interval $(y_0 - \beta, y_0 + \beta)$ and since it also follows from (5) that $\varphi(y_0 - \beta) < 0$ and $\varphi(y_0 + \beta) > 0$, therefore a point y^* can be found between $y_0 - \beta$ and $y_0 + \beta$ for which $\varphi(y^*) = 0$ or, which is the same

$$F(x, y^*) = 0.$$

The number y^* which is uniquely defined in the way described for every x in the interval Δ is a function of x in that interval and we shall denote it by $f(x)$. We have thus proved that *for every x in the interval Δ only one value of y exists which is confined between $y_0 - \beta$ and $y_0 + \beta$ and satisfies the equation (3)*. We have denoted this value by $f(x)$. We have $f(x_0) = y_0$, since the value $y = y_0$ satisfies both the necessary conditions for $x = x_0$.

2°. *Continuity of the function $f(x)$.*

We will now show that the function $f(x)$ which we defined above in the interval Δ is continuous in that interval. Let x_1 be an arbitrary point in the interval Δ and let $\varepsilon > 0$ be as small as we please. Let us assume that $f(x_1) = y_1$ so that $y_0 - \beta < y_1 < y_0 + \beta$. Since the point (x_1, y_1) belongs to the rectangle R' , another rectangle R'' ($x_1 - \lambda \leq x \leq x_1 + \lambda, y_1 - \mu \leq y \leq y_1 + \mu$) can be found with centre at (x_1, y_1) , which lies entirely within R' ; evidently we also have the justification to assume that $\mu < \varepsilon$. We have $F'_y(x, y) > 0$ at every point of the rectangle R'' ; also $F'_y(x_1, y) > 0$ for $|y - y_1| \leq \mu$, i.e. $F'(x_1, y)$ is an increasing function of y in the interval $(y_1 - \mu, y_1 + \mu)$; and since $F(x_1, y_1) = 0$, therefore

$$F(x_1, y_1 - \mu) < 0, F(x_1, y_1 + \mu) > 0.$$

Applying the lemma § 23 again we can conclude that the inequalities

$$F(x, y_1 - \mu) < 0, \quad F(x, y_1 + \mu) > 0$$

hold for all values of x in an interval $(x_1 - \delta, x_1 + \delta)$ and we can assume that $\delta < \lambda$. Therefore for every x in the interval $(x_1 - \delta, x_1 + \delta)$ a number y^* ($y_1 - \mu < y^* < y_1 + \mu$) can be found such that $F(x, y^*) = 0$. Since μ is less than ε , therefore y^* lies in the interval $(y_1 - \varepsilon, y_1 + \varepsilon)$. On the other hand, since $R'' \subset R'$, therefore the interval $(y_1 - \mu, y_1 + \mu)$ and the number y^* lie in the interval $(y_0 - \beta, y_0 + \beta)$. But we have shown in 1° that only one number $y = f(x)$ lies in that interval, which satisfies the equation $F(x, y) = 0$. We therefore have $y^* = f(x)$ and consequently

$$y_1 - \varepsilon < f(x) < y_1 + \varepsilon$$

for every x in the interval $(x_1 - \delta, x_1 + \delta)$. Since ε is arbitrarily small and x_1 an arbitrary point in the interval Δ , it means that the function $f(x)$ is continuous in the interval Δ .

3°. *Uniqueness of the function $f(x)$.*

Let $\varphi(x)$ be a continuous function in the interval Δ so that

$$\varphi(x_0) = y_0, \quad F[x, \varphi(x)] = 0 \quad (x_0 - \alpha \leq x \leq x_0 + \alpha).$$

If we have $y_0 - \beta \leq \varphi(x) \leq y_0 + \beta$ in the interval Δ , then as a result of 1° $\varphi(x)$ is identically equal to $f(x)$. It therefore remains to show that $\varphi(x)$ cannot assume values in the interval Δ , which lie outside the interval $(y_0 - \beta, y_0 + \beta)$. Let $\varphi(x)$ be such a value and let $\varphi(x) > y_0 + \beta$. Since $\varphi(x_0) = y_0 < y_0 + \beta$, therefore it follows from

continuity of the function $\varphi(x)$ that a point x^* can be found between x_0 and $\overline{(x)}$ at which $\varphi(x^*) = y_0 + \beta$. But in that case $F[x^*, \varphi(x^*)] = F[x^*, y_0 + \beta] = 0$, which contradicts the second of the inequalities, (5) since the point x^* evidently belongs to the interval Δ .

4°. *Existence and continuity of $f'(x)$.*

Since, according to our assumption, the function $F(x, y)$ has continuous partial derivatives in the rectangle R and hence also in the rectangle R' , therefore, according to theorem 2 § 90, it is differentiable at every point (x, y) of this rectangle, i.e. in transition from the point (x, y) to the point $(x + \Delta x, y + \Delta y)$ we have:

$$\begin{aligned}\Delta F &= F(x + \Delta x, y + \Delta y) - F(x, y) \\ &= \frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y + o(\rho),\end{aligned}$$

where $\rho = \sqrt{\Delta x^2 + \Delta y^2}$. The increments Δx and Δy can in this case be arbitrary. Assuming that the points x and $x + \Delta x$ lie in the interval $(x_0 - \alpha, x_0 + \alpha)$ we can now assume that

$$y = f(x), y + \Delta y = f(x + \Delta x),$$

so that

$$\Delta y = f(x + \Delta x) - f(x),$$

where Δx remains arbitrary. We evidently have

$$F(x, y) = F(x + \Delta x, y + \Delta y) = 0,$$

and, consequently, also

$$\Delta F = \frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y + o(\rho) = 0,$$

hence

$$\frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y = o(\rho) = o(\sqrt{\Delta x^2 + \Delta y^2}),$$

or

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \frac{\Delta y}{\Delta x} = o\left(\sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}\right) = o\left(1 + \left|\frac{\Delta y}{\Delta x}\right|\right),$$

since we always have $\sqrt{1 + a^2} \leq 1 + |a|$ which can be proved by squaring both sides. It therefore follows that

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \frac{\Delta y}{\Delta x} = \lambda \left(1 + \left|\frac{\Delta y}{\Delta x}\right|\right) = \lambda \left(1 \pm \frac{\Delta y}{\Delta x}\right),$$

$\lambda \rightarrow 0$ for $\varphi \rightarrow 0$; therefore

$$\frac{\Delta y}{\Delta x} = - \frac{\frac{\partial F}{\partial x} - \lambda}{\frac{\partial F}{\partial y} \mp \lambda}.$$

We have shown in 2° that the function $y = f(x)$ is continuous in the interval $(x_0 - \alpha, x_0 + \alpha)$; therefore $\Delta y \rightarrow 0$ as $\Delta x \rightarrow 0$ and therefore also $\rho \rightarrow 0$ which implies in its turn that $\lambda \rightarrow 0$. But the quantities $\partial F / \partial x$ and $\partial F / \partial y$ are constant for $\Delta x \rightarrow 0$ and the latter is non-zero; hence the last equation gives:

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = f'(x) = - \frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}.$$

This evidently concludes the proof of all statements expressed by theorem 1. The expression obtained for $f'(x)$ is the same as what we have obtained in § 92.

It must be stressed that theorem 1, like many other theorems on existence of implicit functions, has a local character: it describes the behaviour of the function F only in a neighbourhood of the point (x_0, y_0) (in the rectangle R) and the behaviour of the function f only in a neighbourhood of the point x_0 . Generally speaking either neighbourhood can be as small as we please.

A natural generalisation of theorem 1 is expressed by the following determining equation:

$$F(x_1, x_2, \dots, x_n, y) = 0 \quad (6)$$

which can have any number of independent variables. If $F = 0$ and $\partial F / \partial y \neq 0$ at the point $(x_{10}, x_{20}, \dots, x_{n0}, y_0)$, then in a neighbourhood of the point $N(x_{10}, x_{20}, \dots, x_{n0})$ a unique continuous function $y = f(x_1, x_2, \dots, x_n)$ exists which satisfies the equation (6) and becomes y_0 at the point N ; this function has continuous partial derivatives with respect to all variables, whose expressions can be readily found. The proof of all these statements is exactly the same as the above proof of theorem 1.

For exercises to § 103 cf. Problem Book by B.P. Demidovich, Section VI, Nos. 232, 235, 237, 275.

§ 104. The general problem

We shall now consider the general problem and restrict ourselves to the consideration of two determining equations. The transition from two equations to three, then from three to four, and so on does not involve further technical difficulties but merely requires a more complicated notation.

Let the following system of equations be given

$$\left. \begin{aligned} F_1(x, y, z) &= 0, \\ F_2(x, y, z) &= 0, \end{aligned} \right\} \quad (1)$$

(for the sake of simplicity of notation we have chosen a problem with one independent variable), where the functions F_1 and F_2 are continuous and have continuous partial derivatives with respect to all the variables in a region P of values of these variables; this region may be chosen in the form of a rectangular parallelepiped. Let $M(x_0, y_0, z_0)$ be an interior point in the region P and let its coordinates satisfy the equation (1):

$$F_1(x_0, y_0, z_0) = 0,$$

$$F_2(x_0, y_0, z_0) = 0.$$

We want to establish the conditions when a unique pair of continuous functions exists

$$y = f_1(x), \quad z = f_2(x),$$

which identically satisfy the equations (1) in a neighbourhood of the point x_0 so that

$$f_1(x_0) = y_0, \quad f_2(x_0) = z_0.$$

We shall also be interested in differentiability of these functions.

Let us at first assume that out of two partial derivatives $\partial F_1 / \partial z$ $\partial F_2 / \partial z$ at least one is non-zero at the point M . Let us assume that

$$\frac{\partial F_1}{\partial z} \neq 0$$

at the point M . It then follows from theorem 1 § 103 (extended to two independent variables) that in a neighbourhood Q of the point $N(x_0, y_0)$ in the (x, y) -plane the unique continuous function

$$z = f(x, y),$$

exists, which satisfies the first of the equation (1) and assumes the value z_0 at the point N ; this function has continuous partial derivatives with respect to x and y in the neighbourhood Q of the point N . Hence in the neighbourhood Q of the point N we have identically with respect to x and y

$$F_1 [x, y, f(x, y)] = 0. \quad (2)$$

Let us now assume

$$F_2 [x, y, f(x, y)] = \Phi(x, y)$$

and note the following. If we succeed in finding a continuous function

$$y = f_1(x),$$

which identically satisfies the equation $\Phi(x, y) = 0$ in a neighbourhood of the point x_0 and is equal to y_0 at the point x_0 , then assuming

$$z = f[x, f_1(x)] = f_2(x), \quad (4)$$

we note directly that the functions f_1 and f_2 satisfy all the necessary requirements; in fact, the relation (2) is identically satisfied with respect to y and therefore remains valid when y is replaced by any continuous function of x , for example, when $y = f_1(x)$ is replaced only by $f_1(x_0) = y_0$; therefore in a neighbourhood of the point x_0 we have identically

$$F_1 \{x, f_1(x), f[x, f_1(x)]\} = F_1 \{x, f_1(x), f_2(x)\} = 0.$$

On the other hand, we have defined the function $y = f_1(x)$ as the solution of the equation

$$\Phi(x, y) = 0,$$

and therefore the relation (3) gives identically in a neighbourhood of the point x_0

$$\begin{aligned} \Phi[x, f_1(x)] &= F_2 \{x, f_1(x), f[x, f_1(x)]\} = \\ &= F_2 \{x, f_1(x), f_2(x)\} = 0. \end{aligned}$$

Hence the functions $y = f_1(x)$, $z = f_2(x)$, in fact, satisfy both equations (1) in a neighbourhood of the point x_0 . On the other hand, it follows from the definition of the function $f_1(x)$ that

$$f_1(x_0) = y_0,$$

and therefore

$$f_2(x_0) = f[x_0, f_1(x_0)] = f(x_0, y_0) = z_0$$

according to the definition of the function f .

Hence we must solve the equation

$$\Phi(x, y) = 0$$

with respect to y . It is evident that all assumptions made with regard to existence of the required solution $y = f_1(x)$ are apparent from theorem 1 § 103 if we have at the $N(x_0, y_0)$:

$$\frac{\partial \Phi}{\partial y} \neq 0;$$

let us now consider what these requirements involve. It follows from definition (3) for the function Φ

$$\frac{\partial \Phi}{\partial y} = \frac{\partial F_2}{\partial y} + \frac{\partial F_2}{\partial z} \frac{\partial f}{\partial y}; \quad (5)$$

but the identity (2) gives us after differentiation with respect to y :

$$\frac{\partial F_1}{\partial y} + \frac{\partial F_1}{\partial z} \frac{\partial f}{\partial y} = 0,$$

hence (since $\partial F_1 / \partial z \neq 0$)

$$\frac{\partial f}{\partial y} = - \frac{\frac{\partial F_1}{\partial y}}{\frac{\partial F_1}{\partial z}};$$

substituting this expression on the right-hand side of the equation (5) we obtain:

$$\begin{aligned} \frac{\partial \Phi}{\partial y} &= \frac{1}{\frac{\partial F_1}{\partial z}} \left[\frac{\partial F_1}{\partial z} \frac{\partial F_2}{\partial y} - \frac{\partial F_1}{\partial y} \frac{\partial F_2}{\partial z} \right] = \\ &= \frac{1}{\frac{\partial F_1}{\partial z}} \begin{vmatrix} \frac{\partial F_1}{\partial z} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial z} & \frac{\partial F_2}{\partial y} \end{vmatrix}. \end{aligned}$$

The condition that $\partial \Phi / \partial y \neq 0$ at the point N is thus equivalent to the requirement that we should have at the point $M(x_0, y_0, z_0)$:

$$\begin{vmatrix} \frac{\partial F_1}{\partial z} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial z} & \frac{\partial F_2}{\partial y} \end{vmatrix} \neq 0.$$

The determinant on the left-hand side of this inequality is known as *Ostrogradskij's determinant* for the functions F_1, F_2 with respect to the variables z, y and we shall denote it by

$$\mathcal{J} = \frac{D(F_1, F_2)}{D(z, y)} *).$$

We thus arrive at the condition that the determinant \mathcal{J} should be non-zero at the point $M(x_0, y_0, z_0)$. Let us note that the condition made from the beginning that out of the two derivatives $\partial F_1/\partial z$ and $\partial F_2/\partial z$ at least one should be non-zero at the point M , simply follows from our new condition since, if both these derivatives vanish, Ostrogradskij's determinant has a column consisting of zeros and it therefore also vanishes.

Let us therefore assume that $\mathcal{J} \neq 0$ at the point M . We then have $\partial\Phi/\partial y \neq 0$ at the point (x_0, y_0) and the solution of our problem is assured. And since the functions $f_1(x)$ and $f_2(x)$ were obtained in this process as a result of the second application of theorem 1 § 103 which always guarantees continuity of the resulting solutions and that of their derivatives, therefore the functions f_1 and f_2 and their derivatives f_1' and f_2' will also be continuous in a neighbourhood of the point x_0 .

We must now prove uniqueness of the solution. Let us assume that we have two functions $f^*_1(x)$ and $f^*_2(x)$ in a neighbourhood of the point x_0 , which are continuous and satisfy the conditions

$$F_1(x, f^*_1, f^*_2) = 0, \quad F_2(x, f^*_1, f^*_2) = 0 \quad (6)$$

and are such that

$$f^*_1(x_0) = y_0, \quad f^*_2(x_0) = z_0; \quad (7)$$

we will show that we have identically $f^*_1 = f_1, f^*_2 = f_2$ in a neighbourhood of the point x_0 .

*) This notation which is generally accepted is connected with the name of the German mathematician Jacobi whom the development of the theory and applications of Ostrogradskij's determinants ("Jacobians") is usually attributed. However, Ostrogradskij himself obtained most of the more important results several years before Jacobi.

We have already mentioned above that the relation (2) is identically satisfied with respect to x and y in a neighbourhood of the point (x_0, y_0) and it will therefore also be satisfied in a neighbourhood of the point x_0 if we replace y by an arbitrary continuous function of x which becomes y_0 at the point x_0 . It follows from (7) that we can take the function $f^*_{-1}(x)$ for this purpose so that

$$F_1[x, f^*_{-1}, f(x, f^*_{-1})] = 0 \quad (8)$$

in a neighbourhood of the point x_0 . But, on the other hand, it follows from (6) that

$$F_1[(x, f^*_{-1}, f^*_{-2})] = 0, \quad (9)$$

and since $z = f(x, y)$ is the *unique* continuous solution of the equation

$$F_1(x, y, z) = 0,$$

which is equal to z_0 at the point (x_0, y_0) , therefore it follows from (8) and (9) that

$$f(x, f^*_{-1}) = f^*_{-2} \quad (10)$$

in a neighbourhood of the point x_0 . But it follows from (3), (10) and (6) that in a neighbourhood of the same point x_0

$$\Phi(x, f^*_{-1}) = F_2[x, f^*_{-1}, f(x, f^*_{-1})] = F_2[x, f^*_{-1}, f^*_{-2}] = 0. \quad (11)$$

And since, by definition, the function $y = f_1(x)$ is the *unique* continuous solution of the equation

$$\Phi(x, y) = 0,$$

which becomes y_0 at the point x_0 , therefore it follows from (11) that

$$f^*_{-1}(x) = f_1(x)$$

in a neighbourhood of the point x_0 ; it also follows from (4) and (10) that

$$f^*_{-2}(x) = f_2(x),$$

which was to be proved.

The result of the above investigation can be stated in the form of the following theorem.

Theorem 1. *Let the functions $F_1(x, y, z)$ and $F_2(x, y, z)$ be continuous and have partial derivatives with respect to all variables in a neighbourhood of the point (x_0, y_0, z_0) and let at that point*

$$F_1 = 0, F_2 = 0, \mathcal{J} = \frac{D(F_1, F_2)}{D(y, z)} \neq 0.$$

In that case a unique pair of continuous functions

$$y = f_1(x), z = f_2(x),$$

exists in the neighbourhood of the point x_0 , which satisfy the equations (1) and the conditions $f_1(x_0) = y_0, f_2(x_0) = z_0$. These functions have continuous derivatives in the neighbourhood of the point x_0 .

We have already mentioned above that we can extend this theorem by means of induction to hold for an arbitrary number m of determining equations with left-hand sides F_1, F_2, \dots, F_m and $m+n$ unknowns $y_1, y_2, \dots, y_m, x_1, x_2, \dots, x_n$. The necessary condition for expressing the variables y_1, y_2, \dots, y_m uniquely as functions of x_1, x_2, \dots, x_n in a neighbourhood of a given point is that at that point

$$\mathcal{J} = \frac{D(F_1, F_2, \dots, F_m)}{D(y_1, y_2, \dots, y_m)} \neq 0,$$

where

$$\mathcal{J} = \begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} & \cdots & \frac{\partial F_2}{\partial y_m} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial F_m}{\partial y_1} & \frac{\partial F_m}{\partial y_2} & \cdots & \frac{\partial F_m}{\partial y_m} \end{vmatrix}$$

is Ostrogradskij's determinant for the system of functions F_1, F_2, \dots, F_m with respect to the variables y_1, y_2, \dots, y_m (here we no longer take notice of the usual conditions of continuity and differentiability which are the same in all cases).

It is very important to note that all propositions established above are of *local* character: in all cases the properties of equations in a neighbourhood of a given point imply existence of solutions in the neighbourhood of a definite point; our theorems, however, tell us nothing of the dimensions of this neighbourhood.

It now remains to show in what way it is possible to express in the general case the derivatives of the required functions f_i in terms

of the partial derivatives of the given functions F_i . Let us consider this problem in relation to the conditions of theorem 1. Since in a neighbourhood of the point x_0 we have identically:

$$\begin{aligned} F_1 [x, f_1(x), f_2(x)] &= 0, \\ F_2 [x, f_1(x), f_2(x)] &= 0, \end{aligned}$$

therefore by differentiating these identities with respect to x we obtain (remembering that we have already established existence and continuity of the derivatives $f_1'(x)$ and $f_2'(x)$)

$$\left. \begin{aligned} \frac{\partial F_1}{\partial x} + \frac{\partial F_1}{\partial y} \frac{df_1}{dx} + \frac{\partial F_1}{\partial z} \frac{df_2}{dx} &= 0, \\ \frac{\partial F_2}{\partial x} + \frac{\partial F_2}{\partial y} \frac{df_1}{dx} + \frac{\partial F_2}{\partial z} \frac{df_2}{dx} &= 0. \end{aligned} \right\} \quad (12)$$

Regarding df_1 / dx and df_2 / dx as the unknowns of this system we obtain for them unique expressions in terms of the partial derivatives of the functions F_1 and F_2 since the determinant of the system (12) is Ostrogradskij's determinant \mathcal{J} which is non-zero in the neighbourhood in which we are interested. We therefore obtain:

$$\begin{aligned} \frac{df_1}{dx} &= \frac{1}{\mathcal{J}} \left\{ \frac{\partial F_2}{\partial x} \frac{\partial F_1}{\partial z} - \frac{\partial F_1}{\partial x} \frac{\partial F_2}{\partial z} \right\}, \\ \frac{df_2}{dx} &= \frac{1}{\mathcal{J}} \left\{ \frac{\partial F_1}{\partial x} \frac{\partial F_2}{\partial y} - \frac{\partial F_2}{\partial x} \frac{\partial F_1}{\partial y} \right\}. \end{aligned}$$

§ 105. Ostrogradskij's determinant

1. General properties. As we have learnt in the last paragraph, the determinant $\mathcal{J} = D(F_1, F_2, \dots, F_m) / D(y_1, y_2, \dots, y_m)$ of the system of m functions F_1, F_2, \dots, F_m of m variables y_1, y_2, \dots, y_m on which these functions depend is of great importance in functional solution of a system of equations (or, which is the same in the theory of implicit functions.) However, the above determinant is also used in other analytical problems and has many applications; we must, therefore, regard this determinant as an important instrument in analytical arguments and calculations. If we are given m functions F_1, F_2, \dots, F_m of m variables y_1, y_2, \dots, y_m and if, within the scope of the problem, we wish to generalise the concept of derivative of one variable to include our case so that this generalised form could be expressed by a *single number* (irrespective of m), then in a majority of cases it is convenient to take the determinant \mathcal{J} as this number.

This was the position in § 104 where we considered existence of implicit functions; this can be clearly seen by comparing the theorems in § 103 and § 104.

The reason behind this general phenomenon which makes us regard the determinant J as a “derivative of the system of functions F_1, F_2, \dots, F_m of a system of variables y_1, y_2, \dots, y_m ” is due to the fact that most important properties of these determinants are similar to those of the usual derivatives; we shall now consider several simple properties but for the sake of simplicity we shall restrict ourselves to the consideration of the case when $m=2$ (although all properties which we shall consider hold for every m).

Let the functions $x = x(u, v)$, $y = y(u, v)$ be continuous and have continuous partial derivatives with respect to u and v within a region of values of these variables and let in this region

$$J = \frac{D(x, y)}{D(u, v)} \neq 0.$$

For the system of functions

$$F_1(x, y, u, v) = x - x(u, v),$$

$$F_2(x, y, u, v) = y - y(u, v),$$

all conditions of theorem 1 § 104 will be satisfied in the neighbourhood of every point (x_0, y_0, u_0, v_0) (where (u_0, v_0) belongs to the given region, $x_0 = x(u_0, v_0)$ and $y_0 = y(u_0, v_0)$). This theorem therefore enables us to conclude that a unique *pair of reciprocal functions*

$$u = u(x, y), \quad v = v(x, y)$$

exists in a neighbourhood of the point (x_0, y_0) ; we can also maintain that these functions are continuous and that their partial derivatives with respect to x and y are continuous.

Let us now assume that we are again given $x = x(u, v)$, $y = y(u, v)$ and let u and v , in their turn, be functions of new variables s and t :

$$u = u(s, t), \quad v = v(s, t).$$

where these functions are subject to the usual conditions of continuity and differentiability. x and y are now “composite” functions of s and t :

$$x = x[u(s, t), v(s, t)], \quad y = y[u(s, t), v(s, t)].$$

According to the rule for differentiating composite functions (§ 92) we have:

$$\frac{\partial x}{\partial s} = \frac{\partial x}{\partial u} \frac{\partial u}{\partial s} + \frac{\partial x}{\partial v} \frac{\partial v}{\partial s}, \quad \frac{\partial y}{\partial s} = \frac{\partial y}{\partial u} \frac{\partial u}{\partial s} + \frac{\partial y}{\partial v} \frac{\partial v}{\partial s},$$

and we have similar formulae for $\partial x/\partial t$ and $\partial y/\partial t$. Therefore

$$\frac{D(x, y)}{D(s, t)} = \begin{vmatrix} \frac{\partial x}{\partial u} \frac{\partial u}{\partial s} + \frac{\partial x}{\partial v} \frac{\partial v}{\partial s} & \frac{\partial y}{\partial u} \frac{\partial u}{\partial s} + \frac{\partial y}{\partial v} \frac{\partial v}{\partial s} \\ \frac{\partial x}{\partial u} \frac{\partial u}{\partial t} + \frac{\partial x}{\partial v} \frac{\partial v}{\partial t} & \frac{\partial y}{\partial u} \frac{\partial u}{\partial t} + \frac{\partial y}{\partial v} \frac{\partial v}{\partial t} \end{vmatrix}. \quad (1)$$

But on the other hand

$$\frac{D(x, y)}{D(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}, \quad (2)$$

$$\frac{D(u, v)}{D(s, t)} = \begin{vmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{vmatrix}. \quad (3)$$

According to a well-known law for multiplying determinants the determinant (1) is equal to the product of the determinants (2) and (3); therefore

$$\frac{D(x, y)}{D(s, t)} = \frac{D(x, y)}{D(u, v)} \cdot \frac{D(u, v)}{D(s, t)}. \quad (4)$$

This relation (which remains valid for determinants with any number m of rows and columns) tells us how to construct Ostrogradskij's determinant for a system of composite functions; this method is exactly similar to the law for differentiating a composite function of one independent variable

$$x = x(u), \quad u = u(s); \quad \frac{dx}{ds} = \frac{dx}{du} \frac{du}{ds}.$$

In particular, assuming that $s = x, t = y$ (i.e. returning from the new variables u, v to the old variables x, y) we obtain from the relation (4)

$$\frac{D(x, y)}{D(u, v)} \cdot \frac{D(u, v)}{D(x, y)} = \frac{D(x, y)}{D(x, y)} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1;$$

this shows that Ostrogradskij's determinants for the system of given functions and for the system of reciprocal functions are mutually reciprocal; this rule is exactly the same as the rule

$$y = y(x); \quad x = x(y); \quad \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

for differentiating reciprocal functions when we have a function of one independent variable.

2. Ostrogradskij's determinant as the local coefficient of an expanding area. We shall now consider one very important geometrical application of Ostrogradskij's determinant which we shall find useful later; here we shall again restrict ourselves to the consideration of the case when $m = 2$.

Let the functions

$$u = u(x, y), \quad v = v(x, y) \quad (5)$$

be continuous and have continuous partial derivatives in a region of the XY -plane and let in this region

$$\frac{D(u, v)}{D(x, y)} \neq 0.$$

Let us analyse the transformations (5) of variables from a geometrical point of view. Let u and v be the cartesian coordinates

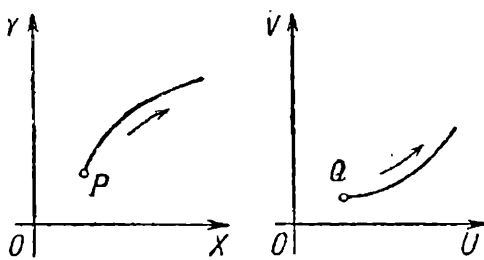


Fig. 63.

of the point (u, v) in a new plane which we shall call the UV -plane. It then follows from the relation (5) that each point $P(x, y)$ in the given XY -plane corresponds to a point $Q(u, v)$ in the UV -plane (fig. 63). If we assume that the point P moves within the region of the XY -plane,

then the corresponding point in the UV -plane will change its position in a quite definite manner. Hence every curve in the XY -plane will correspond to a definite curve in the UV -plane and each figure in the XY -plane to a certain figure in the UV -plane. Also the region in which the functions (5) are defined in the XY plane will be transformed into a new region in the UV -plane.

Let $A(a, b)$ be a definite point in the XY -plane, which lies within the region in which the functions (5) are defined and let h be a small positive number. The points $A(a, b)$, $B(a + h, b)$, $C(a + h, b + h)$, $D(a, b + h)$ (fig. 64, a) are evidently the vertices of a square with side h . As result of the transformation (5) these points become the points A' , B' , C' , D' respectively in the UV -plane (fig. 64, b) whose coordinates are evidently equal to

$$\left. \begin{aligned} A' [u(a, b), v(a, b)], \\ B' [u(a + h, b), v(a + h, b)], \\ C' [u(a + h, b + h), v(a + h, b + h)], \\ D' [u(a, b + h), v(a, b + h)]. \end{aligned} \right\} \quad (6)$$

The square as a whole is transformed into a curvilinear quadrilateral $A'B'C'D'$ which is represented in fig. 64, b . We shall try to find

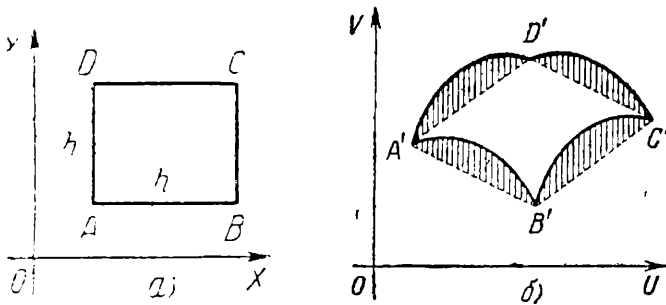


Fig. 63

the approximate value of the area of this curvilinear quadrilateral on the assumption that the number h is very small. We shall at first replace the arcs $A'B'$, $B'C'$, $C'D'$, $D'A'$ of the curves by rectilinear chords, keeping their ends the same, and calculate the area of the *rectilinear* quadrilateral $A'B'C'D'$ (fig. 64, b) which will be equal to the sum of the triangles $A'B'C'$ and $A'D'C'$. According to a well-known formula of analytical geometry the area of a triangle with vertices (u_1, v_1) , (u_2, v_2) , (u_3, v_3) is equal to half of the absolute value of the determinant

$$\begin{vmatrix} u_1 & v_1 & 1 \\ u_2 & v_2 & 1 \\ u_3 & v_3 & 1 \end{vmatrix} = \begin{vmatrix} u_2 - u_1 & v_2 - v_1 \\ u_3 - u_1 & v_3 - v_1 \end{vmatrix}.$$

Hence we obtain the following expression for the area of the triangle $A'B'C'$:

$$\begin{aligned}
& \pm \frac{1}{2} \begin{vmatrix} u(a+h, b) - u(a, b) & v(a+h, b) - v(a, b) \\ u(a+h, b+h) - u(a+h, b) & v(a+h, b+h) - v(a+h, b) \end{vmatrix} = \\
& = \pm \frac{1}{2} \begin{vmatrix} u'_x(a+\theta_1 h, b) h & v'_x(a+\theta_2 h, b) h \\ u'_y(a+h, b+\theta_3 h) h & v'_y(a+h, b+\theta_4 h) h \end{vmatrix} = \\
& = \pm \frac{h^2}{2} \begin{vmatrix} u'_x(a+\theta_1 h, b) & v'_x(a+\theta_2 h, b) \\ u'_y(a+h, b+\theta_3 h) & v'_y(a+h, b+\theta_4 h) \end{vmatrix},
\end{aligned}$$

where $\theta_1, \theta_2, \theta_3, \theta_4$ are numbers confined between 0 and 1.

We assume that the partial derivatives of the functions u and v with respect to x and y are continuous and therefore also uniformly continuous in the (closed) region under consideration. It follows from continuity of the partial derivatives at the point (a, b) that, provided h is small, all elements of the above determinant will differ by as little as we please from the values of the corresponding derivatives at the point (a, b) and the determinant itself will differ by as little as we please from the determinant

$$J(a, b) = \begin{vmatrix} u'_x(a, b) & v'_x(a, b) \\ u'_y(a, b) & v'_y(a, b) \end{vmatrix},$$

i.e. from Ostrogradskij's determinant for the functions u, v of the variables x, y at the point (a, b) . Therefore we have the following expression for the area of the triangle $A'B'C'$ when $h \rightarrow 0$:

$$\frac{h^2}{2} \left\{ |J(a, b)| + o(1) \right\} = \frac{h^2}{2} |J(a, b)| + o(h^2);$$

but a similar calculation shows that the same expression will be obtained for the area of the triangle $A'D'C'$ so that when $h \rightarrow 0$, the area of the whole rectilinear quadrilateral $A'B'C'D'$ is equal to

$$h^2 |J(a, b)| + o(h^2);$$

it is important to remember that the evaluation thus obtained holds *uniformly* with respect to all possible positions of the point (a, b) in the region under consideration. *)

*) This means that, when $h \rightarrow 0$, the ratio of the second term to h^2 tends to zero uniformly with respect to a and b in the given region.

We must now return from the rectilinear quadrilateral $A'B'C'D'$ to the curvilinear quadrilateral with the same vertices. However, the difference in areas of these two quadrilaterals evidently does not exceed the sum of the areas of four narrow strips which are shaded in fig. 64. Therefore in order to show that the expression (7) also represents the area of the *curvilinear* quadrilateral $A'B'C'D'$ it is sufficient to prove that the areas of the shaded strips are equal to a quantity of the type $o(h^2)$; the calculation is naturally the same for all four figures; let us perform this calculation, say, for the figure $A'B'$ (fig. 65) and let us, for the sake of brevity, denote the coordinates of the points A' and B' by (u_0, v_0) and $(u_0 + \Delta u, v_0 + \Delta v)$ respectively.

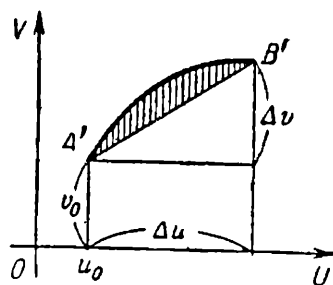


Fig. 65.

Let us assume that $f(a, b) \neq 0$; in that case at least one of the partial derivatives $\partial u / \partial x$, $\partial v / \partial x$ will be non-zero at the point (a, b) ; let $\partial u / \partial x > 0$ at the point (a, b) ; therefore provided h is sufficiently small, we have $\partial u / \partial x > 0$ at every point of the square $ABCD$ and, in particular, at every point on the side AB of this square. Hence if the point (x, y) moves along point that side from A to B , the corresponding (u, v) moves along the curve $A'B'$ in the direction of increasing values of u ; we can therefore represent this curve by an equation in the form $v = f(u)$, where $u_0 \leq u \leq u_0 + \Delta u$. The equation of the rectilinear section $A'B'$ is evidently equal to

$$v = v_0 + \frac{\Delta v}{\Delta u} (u - u_0) = f(u_0) + \frac{f(u_0 + \Delta u) - f(u_0)}{\Delta u} (u - u_0),$$

so that the area S of the figure shaded in fig. 65 can be expressed by the integral

$$S = \int_{u_0}^{u_0 + \Delta u} f(u) - f(u_0) - \frac{u - u_0}{\Delta u} [f(u_0 + \Delta u) - f(u_0)] du.$$

Since y remains constant in the interval AB , therefore u and v become functions of one variable x and we have:

$$du = \frac{\partial u}{\partial x} dx, \quad dv = \frac{\partial v}{\partial x} dx,$$

therefore

$$\frac{dv}{du} = \frac{\frac{\partial v}{\partial x}}{\frac{\partial u}{\partial x}};$$

since $\partial u/\partial x \neq 0$ in the interval AB , it implies existence and continuity of the derivative $f'(u) = dv/du$ in the interval $(u_0, u_0 + \Delta u)$. Hence we have for $u_0 \leq u \leq u_0 + \Delta u$:

$$f(u) - f(u_0) = (u - u_0) f'(u_1), \quad f(u_0 + \Delta u) - f(u_0) = \Delta u f'(u_2),$$

where u_1 and u_2 are confined between u_0 and $u_0 + \Delta u$; thus the expression obtained above for the area S can be written in the form

$$S = \int_{u_0}^{u_0 + \Delta u} (u - u_0) |f'(u_1) - f'(u_2)| du,$$

and since the function $f'(u)$ is continuous for $\Delta u \rightarrow 0$

$$S = o \left(\int_{u_0}^{u_0 + \Delta u} (u - u_0) du \right) = o(\Delta u^2).$$

But

$$\Delta u = u(a + h, b) - u(a, b)$$

is an infinitely small quantity of the same order as h for $h \rightarrow 0$ and we therefore obtain:

$$S = o(h^2),$$

which was to be proved.

We can therefore say that the area of the curvilinear quadrilateral $A'B'C'D'$ resulting from the transformation (5) of the square $ABCD$ is equal to the expression (7); this evaluation holds uniformly in the region under consideration with respect to the position of the point (a, b) . Since the area of the square $ABCD$ is equal to h^2 , the ratio of the transformed and initial areas is equal to

$$|\mathcal{J}(a, b)| + o(1),$$

and the limit of this ratio for $h \rightarrow 0$, is equal to $|\mathcal{J}(a, b)|$.

This result can be greatly generalised. Instead of the square $ABCD$ we could have taken any other sufficiently simple figure containing the point $A(a, b)$ and we could subsequently have shrunk it so that its diameter should tend to zero; in this case (as can be seen from a more detailed analysis which we cannot give here) the ratio of the areas of the transformed and given figures will always

tend to the limit $|\mathcal{J}(a, b)|$. Hence the absolute value of Ostrogradskij's determinant for the transformation (5) can be regarded as the *coefficient of expansion or contraction of areas* resulting from the transformation (5) in the immediate neighbourhood of the point (a, b) . This result is very important in the theory of multiple integrals which we shall study in the next section. The geometrical part played by Ostrogradskij's determinant can be extended to a space of an arbitrary number of dimensions. Thus in the transformation of a three-dimensional space Ostrogradskij's determinant for this transformation gives us the coefficient of volume expansion or contraction of small geometrical bodies which lie completely in the neighbourhood of a given point.

§ 106. Conditional extremum

In this paragraph we shall consider the theory of the so-called *conditional extrema* (maxima and minima) to which theory of implicit functions can be directly applied.

Let a certain surface be given in space, which is defined by the following equation

$$F(x, y, z) = 0 ; \quad (1)$$

we must find a point on this surface at which a function $f(x, y, z)$ assumes the greatest (or smallest) value as compared with other points on that surface. From the analytical point of view this implies finding the maximum (or minimum) of the function $f(x, y, z)$ for all possible combinations of the numbers x, y, z which are connected by the relation (1); this problem differs from the usual extremum problems by the fact that a connecting equation (1) is given, *i.e.* that we are interested in the comparative value of the function $f(x, y, z)$ only at points subjected to the relation (1).

It may happen that the point at which the given function f assumes its greatest or smallest value may be one of the points on a *line* expressed by the equations

$$\left. \begin{aligned} F_1(x, y, z) &= 0, \\ F_2(x, y, z) &= 0. \end{aligned} \right\} \quad (2)$$

From an analytical point of view this implies that among all combinations of the three numbers (x, y, z) which satisfy the relation (2) there must be a point at which the function $f(x, y, z)$ has its greatest or smallest value.

In all such cases we speak of a *conditional extremum* of the function $f(x, y, z)$ bearing in mind the fact that at the point which we are trying to find, the function $f(x, y, z)$ assumes its greatest or smallest value only in relation to points which satisfy some additional *conditions* of the type (1) or (2). These relations are usually called *connecting equations* characteristic of the given problem. The most general type of conditional extremum problem can evidently be formulated as follows: among the points (x_1, x_2, \dots, x_n) which satisfy the connecting equations

$$F_i(x_1, x_2, \dots, x_n) = 0 \quad (i = 1, 2, \dots, m; \quad m < n),$$

a point must exist at which the given function $f(x_1, x_2, \dots, x_n)$ assumes its greatest or smallest value. In such cases, as in ordinary extremum problems, the region of values of the variables x_i ($1 \leq i \leq n$) in which we are interested is usually given beforehand.

If some m variables (for example x_1, x_2, \dots, x_m) can be determined in this region as single valued functions of the other variables (x_{m+1}, \dots, x_n) by means of m connecting equations

$$x_i = \varphi_i(x_{m+1}, x_{m+2}, \dots, x_n) \quad (i = 1, 2, \dots, m),$$

then by substituting their expressions into the formula for the function f we evidently obtain a function of the variables (x_{m+1}, \dots, x_n) whose extremum is now sought among various systems of values of these variables which are no longer interconnected, *i.e.* we now have a simple extremum problem whose solution has been considered in §96. It is therefore obvious that in conditional extremum problems it is very important to solve the system of connecting equations with respect to a certain group of variables; thus the general theory of a conditional extremum is closely related to the theory of implicit functions.

In order to simplify notation and clarify the arguments used we shall now consider the case when $n = 5$, $m = 2$, *i.e.* the problem of the conditional extremum of the function $f(x, y, z, u, v)$ of five variables which are related by two connecting equations:

$$\left. \begin{aligned} F_1(x, y, z, u, v) &= 0, \\ F_2(x, y, z, u, v) &= 0. \end{aligned} \right\} \quad (3)$$

All arguments which we shall use in this connection can be used without modifications for every n and m . Therefore, as in the case of a simple extremum and for the same reasons, we shall only

consider the properties of the *relative (local)* conditional extremum, and, as before, we shall restrict ourselves to the deduction of the *necessary* conditions of general character, for we are now even less able to go into further details.

Let us therefore assume that at a point $M(x_0, y_0, z_0, u_0, v_0)$ whose coordinates satisfy the connecting equation (3) the function $f(x, y, z, u, v)$ assumes the greatest or smallest value as compared with all sufficiently closely situated points whose coordinates also satisfy the equations (3). Let us write a matrix with two rows and five columns

$$\begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} & \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} & \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{pmatrix}$$

and assume that among the determinants of the second order which can be composed of the elements of this matrix there is at least one which is non-zero at the point M . Let us assume that this is the determinant

$$\begin{vmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{vmatrix} = J.$$

In that case (assuming that the usual conditions of continuity and differentiability are satisfied by the functions F_1 and F_2) we can conclude from theorem 1 § 104 that a single pair of functions exists

$$u = u(x, y, z), \quad v = v(x, y, z),$$

which identically satisfies the equations (3) in a neighbourhood of the point $P(x_0, y_0, z_0)$ and for which

$$u(x_0, y_0, z_0) = u_0, \quad v(x_0, y_0, z_0) = v_0.$$

In a neighbourhood of the point P these two functions are continuous and have continuous partial derivatives with respect to all three variables.

Since in our problem we are interested in the values of the function $f(x, y, z, u, v)$ at points close to the point M whose coordinates are connected by the relation (3), we can replace u and v by

the functions $u(x, y, z)$ and $v(x, y, z)$ respectively in the expression of this function and maintain that the function of x, y, z so obtained

$$\varphi(x, y, z) = f[x, y, z, u(x, y, z), v(x, y, z)] \quad (4)$$

has a simple (ordinary) local extremum at the point $P(x_0, y_0, z_0)$. It follows from § 96 that we should therefore have at the point P

$$\frac{\partial \varphi}{\partial x} = \frac{\partial \varphi}{\partial y} = \frac{\partial \varphi}{\partial z} = 0;$$

according to the expression (4) the function φ gives us:

$$\left. \begin{aligned} \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial x} &= 0, \\ \frac{\partial f}{\partial y} + \frac{\partial f}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial y} &= 0, \\ \frac{\partial f}{\partial z} + \frac{\partial f}{\partial u} \frac{\partial u}{\partial z} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial z} &= 0. \end{aligned} \right\} \quad (5)$$

In order that the left-hand sides of these equations could be regarded as given functions of x, y, z, u, v we must express $\partial u / \partial x$, $\partial v / \partial x$, $\partial u / \partial y$, $\partial v / \partial y$, $\partial u / \partial z$, $\partial v / \partial z$ in terms of the five variable functions. But this necessitates differentiation of implicit functions which we have considered in detail in § 104. As before we shall differentiate the following identities with respect to x, y and z :

$$F_1[x, y, z, u(x, y, z), v(x, y, z)] = 0,$$

$$F_2[x, y, z, u(x, y, z), v(x, y, z)] = 0,$$

whence we obtain:

$$\left. \begin{aligned} \frac{\partial F_1}{\partial x} + \frac{\partial F_1}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial F_1}{\partial v} \frac{\partial v}{\partial x} &= 0, \\ \frac{\partial F_2}{\partial x} + \frac{\partial F_2}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial F_2}{\partial v} \frac{\partial v}{\partial x} &= 0 \end{aligned} \right\} \quad (6)$$

and similar relations for the derivatives with respect to y and z . As usual we can use the equation (6) in order to express uniquely the derivatives $\partial u / \partial x$ and $\partial v / \partial x$ in terms of the partial derivatives of the given functions F_1 and F_2 (the determinant of the system (6) is \mathcal{J} and, according to our assumption, it is non-zero at the point M and therefore also in a neighbourhood of that point). It is obvious that systems similar to the system (6) and written for derivatives with

respect to y and z , will give us analogous expressions for $\partial u / \partial y$, $\partial v / \partial y$, $\partial u / \partial z$, $\partial v / \partial z$. Replacing these expressions into the relations (5) we obtain three equations whose left-hand sides can now be regarded as given functions of x, y, z, u, v . Adding to these equations the connecting equations (3) we obtain five equations whose left-hand sides contain known functions of the variables x, y, z, u, v . We have shown above that this system of five equations with five unknowns should give the coordinates of the point M provided this point is the required conditional extremum. It is therefore natural to call every system of five numbers (x, y, z, u, v) , which satisfies the system of equations obtained, as *stationary point* of the given problem. The result can then be stated in the same way as the results obtained earlier for simple extrema: provided the usual conditions of continuity and differentiability of the given functions are preserved, the function can only have local conditional extrema at stationary points. Naturally the question whether this or other stationary point gives the local extremum of the function f and, if so, what is the nature of this extremum, cannot be answered on the basis of the above considerations and must be investigated separately.

We have seen above that the first three equations of a system of five equations available for the determination of stationary points are obtained as a result of solving the system of linear equations (6) with respect to $\partial u / \partial x$ and $\partial v / \partial x$ and the substitution of the resulting solutions into the first of the equations (5); the same operations are subsequently carried out for derivatives with respect to y and z and the results obtained are replaced respectively in the second and third of the equations (5). It is clear that there would be no great difficulty in carrying out all these simple operations and writing the final system of equations in a definite form (this would evidently contain only partial derivatives of the functions f, F_1 and F_2 with respect to all five variables). However, we shall not do so, for in practice we can usually obtain the final system of equations for the determination of stationary points by means of another simple method, *viz* by the so-called "method of undefined factors". We shall now show how this is done.

The sequence of operations described above essentially involves an elementary algebraic operation — elimination of six unknowns $\partial u / \partial x, \partial v / \partial x, \partial u / \partial y, \partial v / \partial y, \partial u / \partial z, \partial v / \partial z$ from nine linear equations ((5), (6) and four analogous equations (6) for the derivatives of the functions u and v with respect to y and z). This elimination can naturally be carried out in different ways. Thus in the method used

above we have obtained expressions for all the unknowns eliminated from the last six equations and substituted these expressions in the equation (5). In practice this method is often inconvenient owing to its asymmetry: the part played by the unknowns u and v differs essentially from that played by the remaining variables x, y, z . The main advantage of the method of undefined factors is due to the equivalence of five variables.

Let us assume as before that the point $M(x_0, y_0, z_0, u_0, v_0)$ gives us the required local conditional extremum of the function f and let us preserve all former assumptions made with regard to the functions f, F_1 and F_2 in the neighbourhood of the point M . The system of equations

$$\left. \begin{aligned} \frac{\partial f}{\partial u} + \lambda_1 \frac{\partial F_1}{\partial u} + \lambda_2 \frac{\partial F_2}{\partial u} &= 0, \\ \frac{\partial f}{\partial v} + \lambda_1 \frac{\partial F_1}{\partial v} + \lambda_2 \frac{\partial F_2}{\partial v} &= 0, \end{aligned} \right\} \quad (7)$$

where all partial derivatives are taken at the point M thus has a single solution (λ_1, λ_2) . Multiplying the equations (6) by λ_1 and λ_2 respectively and adding them term-by-term to the first of the equations (5) we obtain at the point M (as a result of (7)) :

$$\frac{\partial f}{\partial x} + \lambda_1 \frac{\partial F_1}{\partial x} + \lambda_2 \frac{\partial F_2}{\partial x} = 0; \quad (8)$$

if we now write equations analogous to (6) for the derivatives with respect to y and z and combine them in the way described above with the second and third of the equations (5), we evidently obtain, as in the equation (8), the following equations

$$\left. \begin{aligned} \frac{\partial f}{\partial y} + \lambda_1 \frac{\partial F_1}{\partial y} + \lambda_2 \frac{\partial F_2}{\partial y} &= 0, \\ \frac{\partial f}{\partial z} + \lambda_1 \frac{\partial F_1}{\partial z} + \lambda_2 \frac{\partial F_2}{\partial z} &= 0. \end{aligned} \right\} \quad (9)$$

The system of the five equations (7), (8) and (9) is evidently completely symmetrical with respect to the five variables x, y, z, u, v . Adding the connecting equations to these five equations we obtain a system of seven equations with the unknowns $x, y, z, u, v, \lambda_1, \lambda_2$, which (when the usual conditions of continuity and differentiability are preserved) should be satisfied at every stationary point.

If together with the given function f we are also considering the function

$$\Phi = f + \lambda_1 F_1 + \lambda_2 F_2$$

of the same variables, where λ_1 and λ_2 are undefined numerical factors, then the system of equations (7), (8), (9) obtained can be written in the form

$$\frac{\partial \Phi}{\partial x} = \frac{\partial \Phi}{\partial y} = \frac{\partial \Phi}{\partial z} = \frac{\partial \Phi}{\partial u} = \frac{\partial \Phi}{\partial v} = 0,$$

and it therefore represents the system of equations which we would have obtained if instead of the conditional extremum of the function f we would have sought the ordinary extremum of the function Φ . In this reduction the definitions of the conditional stationary points of the function f and finding of the ordinary stationary points of the function Φ involves the practical application of the method of undefined factors. We thus see that the system of equations which determine the conditional stationary points of the function f have two more unknowns (λ_1 and λ_2) than those for ordinary stationary points; we now also have two more equations since the connecting equations are added.

Example. The paraboloid of rotation

$$x^2 + y^2 = z \tag{10}$$

is intersected by the plane

$$x + y + z = 1 \tag{11}$$

in an ellipse; find the greatest and least distances of points on this ellipse from the origin of coordinates.

From an analytical point of view this problem evidently necessitates finding of the maximum and minimum of the function

$$x^2 + y^2 + z^2$$

where the equations (10) and (11) are the connecting equations. Using the method of undefined factors we construct the function

$$\Phi(x, y, z) = x^2 + y^2 + z^2 + \lambda_1(x^2 + y^2 - z) + \lambda_2(x + y + z - 1)$$

and equate to zero its partial derivatives with respect to all three variables; this gives :

$$x = y = -\frac{\lambda_2}{2(\lambda_1 + 1)}, \quad z = \frac{\lambda_1 - \lambda_2}{2};$$

substituting these expressions in the connecting equations (10) and (11) we obtain after easy calculations:

$$\lambda_1 = -3 \pm \frac{5}{3}\sqrt{3}, \quad \lambda_2 = -7 \pm \frac{11}{3}\sqrt{3},$$

and therefore

$$x = y = \frac{-1 \pm \sqrt{3}}{2}, \quad z = 2 \mp \sqrt{3}.$$

This gives two values $9 \mp 5\sqrt{3}$ for the quantity $x^2 + y^2 + z^2$; since the case under consideration existence of the required extrema is geometrically obvious, therefore on further investigations into the nature of the two stationary points obtained are needed and we can consider the problem solved.

The reader will find further exercises in the Problem Book by B.P. Demidovich, Section VI; we recommend Nos. 447, 448, 453, 456, 465.

CHAPTER XXV

GENERALISED INTEGRALS

§ 107. Integrals with infinite limits

In this chapter we shall consider two wider concepts of definite integrals which are very important in the further development of theory and its applications.

The function $y = 1/x^2$ is positive in the region $x \geq 1$; it is continuous, decreases constantly as x increases and tends to zero as $x \rightarrow \infty$. Let us consider the area below the curve $y = 1/x^2$ and above the OX -axis between the abscissae 1 and $b > 1$; we know that this area (fig. 66) can be expressed by the integral

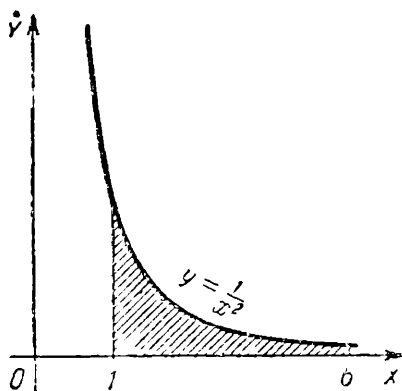


Fig. 66.

$$\int_1^b \frac{dx}{x^2} = 1 - \frac{1}{b}. \quad (1)$$

This area increases as b increases indefinitely, the shaded figure in fig. 66 will extend indefinitely and reach more and more to the right; however, its area remains bounded and tends to unity, as can be seen from the relation (1). This phenomenon is exactly similar to the summation of a convergent series with positive terms, for example, a simple geometrical progression; partial sums of series with an increasing number of terms, thus increasing continuously do not, however increase indefinitely but tend to a definite limit; similarly here the shaded part includes an indefinitely large area as b increases and thus increases continuously but not indefinitely and tends to a definite finite limit. And as before we have

agreed to call the limit of the partial sums as the sum of "all" terms of this series, so now the limit of the shaded figure can be called the area of *whole* figure extending to infinity and bounded from above by the curve $y = 1/x^2$, from below by the OX -axis and from left by the straight line $x = 1$ (it is not bounded at all on the right).

However, partial sums only have limits for convergent series; we know that there exist series whose terms are positive and decrease monotonically tending to zero while the series themselves diverge so that their partial sums increase indefinitely without tending to any limit (for example a harmonic series). We can have a similar position here with the area of the figure which extends to infinity. Thus the curve $y = 1/x$ has a similar course to the curve $y = 1/x^2$ (fig. 66) in the region $x > 1$; but since for $b \rightarrow \infty$

$$\int_1^b \frac{dx}{x} = \ln b \rightarrow \infty,$$

therefore in this case the area of the shaded figure increases *indefinitely* as b increases so that the whole figure extending to infinity has no longer a finite area.

Let us agree in general to call the limit, in case it exists, of every function $f(x)$ integrable in the interval (a, b) where b is as large as we please :

$$\lim_{b \rightarrow \infty} \int_a^b f(x) dx, \quad (2)$$

as the *generalised integral* of the function $f(x)$ in the subinterval $(a, +\infty)$ (or within the limits from a to $+\infty$), and denote this limit by

$$\int_a^{\infty} f(x) dx. \quad (3)$$

If the limit (2) exists, then the integral (3) is said to be *convergent* and the limit (2) is said to be the value of this integral. If the limit (2) does not exist, then the integral (3) is said to be *divergent* and has no value. Here the function $f(x)$ must not necessarily be positive or monotone; the above definition will have a definite meaning provided the function $f(x)$ is integrable for every $b > a$ in the interval (a, b) ; this means, it is sufficient to assume that the function $f(x)$ is

continuous in the region $x \geq a$. In this general case the simple geometrical interpretation of the integral (3) which we have used in the beginning will evidently no longer hold.

We have so far assumed that the lower limit of integration remains constant whereas the upper limit increases indefinitely. We shall evidently have a similar case when the position is reversed, *i.e.*, when the upper limit of integration b remains constant whereas the lower limit of integration a , which is negative, increases indefinitely in its absolute value ($a \rightarrow -\infty$). If the function $f(x)$ is integrable in the interval (a, b) for every $a < b$ and if the following limit exists :

$$\lim_{a \rightarrow -\infty} \int_a^b f(x) dx, \quad (4)$$

we can denote this limit by

$$\int_{-\infty}^b f(x) dx \quad (5)$$

and say that the above integral is convergent and equal to the limit of (4); if this limit does not exist, then the integral (5) is divergent and has no definite value.

Finally the case when $a \rightarrow -\infty$ and $b \rightarrow +\infty$ simultaneously and independent of one another is also possible, *i.e.*, the interval of integration increases and covers the whole number line. We assume in this case that the integral

$$\int_a^b f(x) dx = I(a, b)$$

tends to the limit I and write

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} I(a, b) = I, \quad (6)$$

if for every $\varepsilon > 0$ an $A > 0$ can be found such that for $a < -A$ $b > A$ always

$$|I(a, b) - I| < \varepsilon.$$

Obviously it is necessary and sufficient for the limit (6) to exist that the following two integrals should converge

$$\int_0^{+\infty} f(x) dx \quad \text{and} \quad \int_{-\infty}^0 f(x) dx,$$

and when this is so, the limit (6) is equal to the sum of these two integrals. If the limit (6) exists, we can denote it by

$$\int_{-\infty}^{+\infty} f(x) dx, \quad (7)$$

and we say that the integral (7) converges; otherwise we say that the integral (7) diverges and we are not able to give it a numerical value.

Example 1. Since

$$\int_a^b \frac{dx}{1+x^2} = \arctan b - \arctan a,$$

therefore

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b \frac{dx}{1+x^2} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi;$$

the integral

$$\int_{-\infty}^{+\infty} \frac{dx}{1+x^2}$$

is therefore convergent and equal to π .

Example 2. Since

$$\int_a^b \cos x dx = \sin b - \sin a$$

and since $\sin x$ does not tend to a limit for $x \rightarrow \infty$, therefore, the integrals

$$\int_a^{+\infty} \cos x dx, \quad \int_{-\infty}^b \cos x dx, \quad \int_{-\infty}^{+\infty} \cos x dx$$

are all divergent and have no numerical values.

Example 3. Since

$$\int_a^b e^x dx = e^b - e^a,$$

therefore the integral

$$\int_{-\infty}^b e^x dx$$

is convergent and equal to e^b (since $e^a \rightarrow 0$ as $a \rightarrow -\infty$); on the other hand the integral

$$\int_a^{+\infty} e^x dx$$

is divergent (since $e^b \rightarrow +\infty$ as $b \rightarrow +\infty$). Therefore the following integral is also divergent

$$\int_{-\infty}^{+\infty} e^x dx.$$

The analogy between infinite series and generalised integrals prompts us to seek methods for determining convergence of generalised integrals on the same lines as in finding methods for determining convergence of infinite series. In future we shall only deal with integrals of the form \int_a^∞ ; however, all that we shall deduce can be applied without essential modifications to integrals of the type $\int_{-\infty}^b$ and $\int_{-\infty}^{+\infty}$.

At first the general theorem 2 § 19 gives us the necessary and sufficient test for convergence of the generalised integral

$$\int_a^\infty f(x) dx \tag{8}$$

in the form of the condition that for $\varepsilon > 0$ and for all sufficiently large b_1 and b_2 we must have

$$\left| \int_a^{b_2} f(x) dx - \int_a^{b_1} f(x) dx \right| < \varepsilon ;$$

since

$$\int_a^{b_2} f(x) dx - \int_a^{b_1} f(x) dx = \int_{b_1}^{b_2} f(x) dx,$$

therefore this gives us the following test.

Theorem 1. *In order that the integral (8) should be convergent it is necessary and sufficient that for arbitrarily small $\varepsilon > 0$ the following inequality should hold for all sufficiently large b_1 and b_2 :*

$$\left| \int_{b_1}^{b_2} f(x) dx \right| < \varepsilon$$

(in other words, in order that a generalised integral should converge it is necessary and sufficient that any of its sufficiently far removed (and as far as we please) "part" should be as small as we please).

In the same way as the necessary and sufficient condition for convergence of a series with constant signs is the limit of its partial sums, so in order that the integral (8) with a *non-negative integrand* $f(x)$ should converge it is evidently necessary and sufficient that the integral

$$\int_a^b f(x) dx$$

should remain bounded for $b \rightarrow \infty$. This fact enables us to establish the principle of comparison of integrals which is analogous to the principle of comparison of series with constant signs (theorem 1 § 68).

Theorem 2. *If we have $0 \leq f(x) \leq c \varphi(x)$ for $a \leq x < +\infty$ (where $c > 0$ is a constant) and if the functions $f(x)$ and $\varphi(x)$ are integrable in every interval (a, b) ($a < b$), then convergence of the integral*

$$\int_a^\infty \varphi(x) dx$$

implies convergence of the integral

$$\int_a^{\infty} f(x) dx,$$

and the following inequality also holds :

$$\int_a^{\infty} f(x) dx \leq c \int_a^{\infty} \varphi(x) dx.$$

The proof is similar to that of the principle of comparison of series and we can therefore leave it to the reader.

Example 4. Since α is constant and $x \rightarrow +\infty$, we always have $x^\alpha e^{-\frac{1}{2}x} \rightarrow 0$ (example 7 § 37), and therefore for a sufficiently large x

$$x^\alpha e^{-\frac{1}{2}x} < 1,$$

hence

$$x^\alpha e^{-x} = x^\alpha e^{-\frac{1}{2}x} e^{-\frac{1}{2}x} < e^{-\frac{1}{2}x}.$$

It therefore follows from convergence of the integral

$$\int_1^{\infty} e^{-\frac{1}{2}x} dx$$

that the integral

$$\int_1^{\infty} x^\alpha e^{-x} dx$$

converges for every constant α .

Example 5. $\int_1^{\infty} e^{-x} dx = \frac{1}{e}$ is convergent ; since $e^{-x^2} \leq e^{-x}$ for

$x \geq 1$, the following integral is also convergent :

$$\int_1^{\infty} e^{-x^2} dx,$$

although we do not know its value.

We shall not consider simple properties of generalised integrals which are analogous to the corresponding general properties of infinite series and proved in the same way. Thus it can be readily shown that by changing the function $f(x)$ arbitrarily in a *finite interval* (as long as it remains integrable) we cannot affect convergence of the integral (8) (although we generally change its value). Moreover, if both integrals

$$I_1 = \int_a^{\infty} f_1(x) dx, \quad I_2 = \int_a^{\infty} f_2(x) dx$$

are convergent, then the integral

$$I = \int_a^{\infty} \{ f_1(x) \pm f_2(x) \} dx$$

is also convergent and

$$I = I_1 \pm I_2.$$

The integral (8) is said to be *absolutely convergent* if the following integral converges

$$\int_a^{\infty} |f(x)| dx. \quad (9)$$

Convergence of the integral (9) implies convergence of the integral (8); in fact, if the integral (9) is convergent, it follows from theorem 1 that for arbitrarily small $\varepsilon > 0$ we have for every sufficiently large b_1 and b_2

$$\left| \int_{b_1}^{b_2} |f(x)| dx \right| < \varepsilon.$$

But we know (last theorem of § 51) that

$$\left| \int_{b_1}^{b_2} f(x) dx \right| \leq \left| \int_{b_1}^{b_2} |f(x)| dx \right|;$$

therefore for every sufficiently large b_1 and b_2

$$\left| \int_{b_1}^{b_2} f(x) dx \right| < \varepsilon,$$

and again as a result of theorem 1 this implies convergence of the integral (8).

As in the case of infinite series theorem 2 (the principle of comparison) enables us to establish several concrete tests for convergence of generalised integrals which are convenient in practice; we shall now consider some simple tests of this kind.

Theorem 3. *If $\alpha > 1$ and the following inequality holds for all sufficiently large values of x : $|f(x)| < cx^{-\alpha}$, where $c > 0$ is a constant, then the integral (8) is absolutely convergent; conversely, if $\alpha \leq 1$ and for all sufficiently large values of x we have $f(x) > x^{-\alpha}$, then the integral (8) is divergent.*

We assume, as usual, that the function $f(x)$ is integrable in every finite interval (a, b) ($a < b$).

Since the integral

$$\int_a^{\infty} x^{-\alpha} dx$$

(where $a > 0$) is convergent for $\alpha > 1$ and divergent for $\alpha \leq 1$, therefore theorem 3 follows directly from theorem 2; that we must restrict ourselves only to sufficiently large values of x is, of course, due to the fact that changes in the function $f(x)$ in a finite interval do not affect convergence of the integral (8).

It follows directly from theorem 3 that integrals, like those shown below, must be absolutely convergent

$$\int_1^{\infty} \frac{\sin x}{x^2} dx, \quad \int_0^{\infty} \frac{x dx}{(1+x)^3},$$

etc. The test for convergence established by this theorem has many practical applications; however we must regard it as rather rough, since it can only be used (as can be seen from its formula) to establish

absolute convergence of integrals (and other better tests can be found). We therefore give below another much more sensitive test.

Theorem 4. *If $\alpha > 0$, $a > 0$ and the function $\varphi(x)$ is continuous for every $x \geq a$ and a positive number C exists such that for all $b > a$*

$$\left| \int_a^b \varphi(x) dx \right| < C,$$

then the integral

$$\int_a^{\infty} \frac{\varphi(x)}{x^{\alpha}} dx$$

is convergent.

Proof. Let us assume that

$$\int_a^x \varphi(u) du = \Phi(x),$$

so that

$$|\Phi(x)| < C \quad (a < x < +\infty).$$

Integrating by parts we obtain :

$$\int_a^x \frac{\varphi(u)}{u^{\alpha}} du = \int_a^x \frac{\Phi'(u)}{u^{\alpha}} du = \left(\frac{\Phi(u)}{u^{\alpha}} \right) \Big|_a^x + \alpha \int_a^x \frac{\Phi(u)}{u^{\alpha+1}} du.$$

If we now assume that x increases indefinitely, then the first term on the right-hand side tends to zero so that

$$\left| \frac{\Phi(x)}{x^{\alpha}} \right| < \frac{C}{x^{\alpha}} \rightarrow 0 \quad (x \rightarrow \infty),$$

and, on the other hand, $\Phi(a) = 0$ according to the definition of the function $\Phi(x)$. The second term on the right-hand side tends to the following integral as its limit for $x \rightarrow \infty$:

$$\alpha \int_a^{\infty} \frac{\Phi(u)}{u^{\alpha+1}} du \tag{10}$$

which is (absolutely) convergent in accordance with theorem 3, since $\alpha > 0$ and $|\Phi(u)| < C$. Hence the limit

$$\lim_{x \rightarrow \infty} \int_a^x \frac{\phi(u)}{u^\alpha} du = \int_a^\infty \frac{\phi(u)}{u^\alpha} du$$

exists (and is equal to the value of the absolutely convergent integral (10)).

Theorem 4 can be successfully used in order to establish convergence of many integrals which play an important part in various applications; a typical example of integrals of this kind is the integral

$$\int_0^\infty \frac{\sin x}{x} dx, \quad (11)$$

which, according to theorem 4, is convergent so that

$$\left| \int_0^x \sin u du \right| = |1 - \cos x| \leq 2 \quad (0 < x < +\infty).$$

We will show that the integral (11) is not *absolutely* convergent (or, as is usually said, *conditionally* convergent), i.e. the integral

$$\int_a^\infty \frac{|\sin x|}{x} dx$$

($a > 0$) is divergent. As always, $|\sin x| \geq \sin^2 x$ and it follows from the principle of comparison (theorem 2) that for this purpose it is sufficient to prove that the following integral is divergent :

$$\int_a^\infty \frac{\sin^2 x}{x} dx = \int_a^\infty \frac{1 - \cos 2x}{2x} dx. \quad (12)$$

But the integral

$$\int_a^\infty \frac{\cos 2x}{2x} dx \quad (13)$$

is similar to the integral (11) and its convergence can be readily established by means of theorem 4. Thus if the integral (12) would be convergent, then by adding the integral (13) to it we would obtain a sum

$$\int_a^{\infty} \frac{dx}{2x}$$

which should also be convergent; however, this is not so; therefore the integral (12) is divergent and the integral (11) is only conditionally convergent. We have a similar position with all integrals of the type

$$\int_a^{\infty} \frac{\sin x}{x^{\alpha}} dx, \quad \int_a^{\infty} \frac{\cos x}{x^{\alpha}} dx,$$

if $0 < \alpha \leq 1$.

For exercises *cf.* Problem Book by B. P. Demidovich, Section VI, Nos. 108, 109, 111, 120.

In this paragraph we had many occasions to observe the way in which the analogy between infinite series and generalised integrals is of decisive importance in determining fundamental concepts and elucidating main properties of integrals with infinite limits. We will now show how the concept of generalised integrals can be conversely used for making deductions which are very important in the theory of infinite series; by using the concept of generalised integrals we shall establish a test for convergence of series with constant signs, which by its validity and convenience in practical applications has many advantages over all the elementary tests established in § 68.

Theorem 5. (Cauchy's integral test for convergence of series). *Let $f(x)$ be a positive non-increasing continuous function defined for every $x \geq a$, where a is a constant natural number. In that case the series*

$$f(a) + f(a+1) + \dots + f(a+k) + \dots \quad (14)$$

will be convergent or divergent according as the following integral is convergent or divergent :

$$\int_a^{\infty} f(x) dx. \quad (15)$$

Proof. Since the function $f(x)$ is a non-increasing function; therefore we have $f(a+k) \geq f(x) \geq f(a+k+1)$ for $a+k \leq x \leq a+k+1$ and consequently

$$f(a+k) \geq \int_{a+k}^{a+k+1} f(x) dx \geq f(a+k+1) \quad (k=0, 1, 2, \dots).$$

Summing these inequalities with respect to k from 0 to n we obtain:

$$\sum_{k=0}^n f(a+k) \geq \int_a^{a+n+1} f(x) dx \geq \sum_{k=0}^n f(a+k+1).$$

If the integral (15) is convergent, then the central part in these inequalities remains bounded for $n \rightarrow \infty$; the right-hand side will also be bounded and this implies convergence of the series (14) with constant signs; if, however, the integral (15) is divergent, then the central part will increase indefinitely for $n \rightarrow \infty$; the left-hand side will also increase and this shows that the series (14) is divergent. Theorem 5 is thus proved.

Example 6. In § 68 we have considered an important class of series with constant signs of the form

$$\frac{1}{1^s} + \frac{1}{2^s} + \dots + \frac{1}{n^s} + \dots \quad (16)$$

and proved that the series (16) converges for $s > 1$ and diverges for $s \leq 1$. Theorem 5 can be used to establish convergence of the series (16) directly. Assuming in theorem 5 that $a = 1$, $f(x) = x^{-s}$, we can see that convergence of the series (16) is equivalent to convergence of the integral

$$\int_1^{\infty} x^{-s} dx,$$

which is convergent for $s > 1$ and divergent for $s \leq 1$.

Example 7. Let us now consider a more sensitive problem on convergence of series of the type

$$\sum_{n=2}^{\infty} \frac{1}{n (\ln n)^s}, \quad (17)$$

where s is a constant real number. It can be readily shown that if $s > 0$, none of the tests considered in § 68 can be used to establish convergence of series of this type. However, this problem can be readily solved by means of theorem 5. Let us assume that $a = 2, f(x) = 1/x (\ln x)^s$, so that convergence of the series (17) is equivalent to convergence of the integral

$$\int_2^{\infty} \frac{dx}{x (\ln x)^s}. \quad (18)$$

Since

$$\int_2^x \frac{du}{u (\ln u)^s} = \begin{cases} \frac{1}{1-s} \{(\ln x)^{1-s} - (\ln 2)^{1-s}\} & (s \neq 1), \\ \ln \ln x - \ln \ln 2 & (s = 1), \end{cases}$$

therefore the integral (18) is convergent for $s > 1$ (it is equal to $1/(s-1)(\ln 2)^{s-1}$) and divergent for $s \leq 1$; hence the series (17) is also convergent for $s > 1$ and divergent for $s \leq 1$. The series

$$\sum_{n=2}^{\infty} \frac{1}{n \ln n}$$

is also divergent whereas the series

$$\sum_{n=2}^{\infty} \frac{1}{n \ln^2 n}$$

is convergent.

For further exercises cf. Problem Book by B. P. Demidovich, Section V, No 64.

§ 108. Integrals of unbounded functions

In defining the concept of an integral we have so far always assumed that the integrand is bounded within the interval of integration. We shall now introduce a wider concept of an integral which will enable us in certain cases to integrate unbounded functions. As in § 107 we shall begin by considering a simple case. Let the

function $f(x)$ be defined in the interval $(0, 1)$ as follows :

$$f(x) = \begin{cases} \frac{1}{\sqrt{x}} & (0 < x \leq 1), \\ 0 & (x = 0). \end{cases}$$

Since $1/\sqrt{x}$ increases indefinitely as $x \rightarrow 0$, therefore the function $f(x)$ is not bounded in the interval $(0, 1)$. It is discontinuous at the point $x = 0$ and continuous at all other points in the interval $(0, 1)$. Its graph is shown in fig. 67. It is evident that for arbitrarily small $\varepsilon > 0$, we can integrate the function $f(x)$ in the interval $(\varepsilon, 1)$ and its integral

$$\begin{aligned} \int_{\varepsilon}^1 f(x) dx &= \int_{\varepsilon}^1 \frac{dx}{\sqrt{x}} = (2\sqrt{x}) \Big|_{\varepsilon}^1 = \\ &= 2(1 - \sqrt{\varepsilon}) \end{aligned} \quad (1)$$

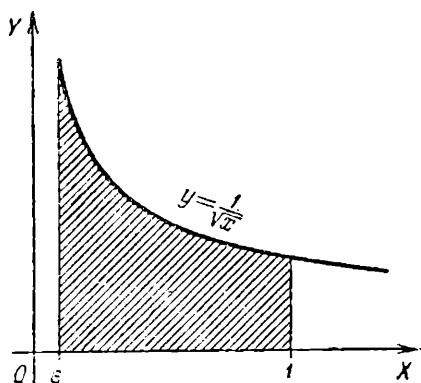


Fig. 67

expresses the area of a curvilinear trapezium which is shaded in fig. 67; this area increases indefinitely as ε decreases; when $\varepsilon \rightarrow 0$, the shaded figure extends indefinitely upwards; however, it can be seen from formula (1) that the area of this figure does not increase indefinitely in this process but merely tends to the limit 2. We naturally take this limit as the area of the whole region above the interval $(0, 1)$ of the OX -axis and below the curve $y = 1/\sqrt{x}$. This geometrical illustration again gives us an example of a figure which has a finite area although it extends to infinity. A comparison of fig. 67 and fig. 66 readily shows the close resemblance of the two pictures.

From a purely analytical point of view we have here a case in which we are unable to determine the above area by means of the integral

$$\int_0^1 f(x) dx,$$

since the integrand is unbounded in the interval $(0, 1)$; but we assume that this area is equal to the limit

$$\lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^1 f(x) dx; \quad (2)$$

for arbitrarily small $\varepsilon > 0$, the integral

$$\int_{\varepsilon}^1 f(x) dx$$

has a definite meaning, since the function $f(x)$ is continuous in the interval of integration.

The limit (2) is called the (generalised) integral of an unbounded function $f(x)$ from 0 to 1 (or in the interval $(0, 1)$) and simply denoted by

$$\int_0^1 f(x) dx;$$

we can therefore write

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^1 \frac{dx}{\sqrt{x}} = 2.$$

Let us now consider the general definition. Let the function $f(x)$ be defined in the interval (a, b) and be integrable for arbitrarily small $\varepsilon > 0$, (and therefore bounded) in the $(a + \varepsilon, b)$ but not bounded in the whole interval (a, b) . If the following limit exists in this case :

$$\lim_{\varepsilon \rightarrow 0} \int_{a+\varepsilon}^b f(x) dx, \quad (3)$$

then we simply call this limit the (generalised) *integral of the unbounded function* $f(x)$ in the interval (a, b) and denote it by

$$\int_a^b f(x) dx; \quad (4)$$

the integral (4) is, in this case, said to be convergent; if, however, no limit exists, then the integral (4) is said to be *divergent* and has no numerical value.

We can thus see that this wider concept of an integral is used in cases when the integrand is only unbounded in the immediate neighbourhood of a point in the interval of integration and bounded and integrable at all other points in this interval. In our example, as well as in the general case, one such "singularity" is the left end a of the interval of integration. It is, however, self-evident that this definition holds for every position of the "singularity" in the interval of integration. Thus if the function $f(x)$ is integrable for every $\varepsilon > 0$ in the interval $(a, b - \varepsilon)$ but not bounded in the whole interval (a, b) and if the integral

$$\int_a^{b-\varepsilon} f(x) dx$$

tends to a limit for $\varepsilon \rightarrow 0$, then, according to the definition, we assume that

$$\int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0} \int_a^{b-\varepsilon} f(x) dx;$$

here the right end b of the interval of integration is a singularity. Finally if an arbitrary interior point c is the singularity in the interval (a, b) , i.e. if the function $f(x)$ is integrable in each of the intervals $(a, c - \varepsilon_1)$ and $(c + \varepsilon_2, b)$ for arbitrarily small $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, then we simply assume that

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx,$$

where both integrals on the right-hand side are generalised integrals with a singularity c at one end of the interval of integration and we can therefore regard them as defined. It is self-evident that in order that the integral

$$\int_a^b f(x) dx$$

should be convergent it is here necessary and sufficient that both the limits

$$\lim_{\varepsilon_1 \rightarrow 0} \int_a^{c - \varepsilon_1} f(x) dx = \int_a^c f(x) dx$$

and

$$\lim_{\varepsilon_2 \rightarrow 0} \int_{c + \varepsilon_2}^b f(x) dx = \int_c^b f(x) dx$$

should exist.

We have already drawn attention to the fact that if the integrand is positive, the geometrical illustration of the integration of an unbounded function resembles the corresponding illustration for integrals with infinite limits as considered in § 107. It can be readily shown that these two problems are closely related to one another analytically. Let us assume, for example, that the function $f(x)$ is unbounded in the neighbourhood of the left end a of the interval of integration so that

$$\int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0} \int_{a + \varepsilon}^b f(x) dx. \quad (5)$$

The transformation of the variable $x = a + \frac{1}{y}$ gives

$$\int_{a + \varepsilon}^b f(x) dx = \int_{\frac{1}{b-a}}^{\frac{1}{\varepsilon}} f\left(a + \frac{1}{y}\right) \frac{dy}{y^2} = \int_{\frac{1}{b-a}}^{\frac{1}{\varepsilon}} \varphi(y) dy,$$

where it is assumed that

$$\varphi(y) = \frac{1}{y^2} f\left(a + \frac{1}{y}\right).$$

The relation (5) therefore gives us :

$$\int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0} \int_{\frac{1}{b-a}}^{\frac{1}{\varepsilon}} \varphi(y) dy = \int_{\frac{1}{b-a}}^{\infty} \varphi(y) dy,$$

so that the integral of an unbounded function can be transformed into an integral with an infinite limit simply by transforming the variable of integration. As a result of this connection between the two new concepts of integrals the properties of integrals of the first type considered in § 107 correspond to analogous properties of integrals of unbounded functions. Hence all fundamental concepts in the theory of integrals of unbounded functions can be constructed on lines parallel to the fundamental concepts in the theory of integrals with infinite limits as considered in § 107. The proofs of all propositions can be carried out equally by either one of two methods: they can be constructed in full analogy with the arguments used in § 107 (which, in their turn, were mostly carried out by analogy to the theory of infinite series) or we can use the above method of transformation of the variable of integration and thus convert the proposition to be proved into the corresponding theorem on integrals with infinite limits and then refer to the appropriate theorem.

Here the following propositions hold, which are analogous to the corresponding theorems considered in § 107 (we assume in all cases that the integrands are bounded and integrable in the interval $(a + \varepsilon, b)$ for arbitrarily small $\varepsilon > 0$ but in general not bounded in the whole interval (a, b)).

Theorem 1'. *In order that the integral (4) should converge it is necessary and sufficient that for arbitrarily small $\varepsilon > 0$ the following inequality should hold for all sufficiently small $\delta_1 > 0$ and $\delta_2 > 0$:*

$$\int_{a + \delta_1}^{a + \delta_2} f(x) dx < \varepsilon.$$

Theorem 2' (Principle of comparison). *If we have $0 \leq f(x) \leq c\varphi(x)$ for $a \leq x \leq b$, where c is a constant positive number, then convergence of the integral*

$$\int_a^b \varphi(x) dx$$

implies convergence of the integral

$$\int_a^b f(x) dx, \tag{4}$$

and the following inequality holds :

$$\int_a^b f(x) dx \leq c \int_a^b \varphi(x) dx.$$

The integral (4) is said to be *absolutely convergent* provided the following integral is convergent :

$$\int_a^b |f(x)| dx; \quad (6)$$

according to theorem 1' convergence of the integral (6) implies convergence of the integral (4).

Since the integral

$$\int_a^b (x-a)^{-\alpha} dx,$$

is convergent for $\alpha < 1$ and divergent for $\alpha \geq 1$, as can be easily shown, the following simple test for convergence can be deduced from theorem 2'.

Theorem 3'. *If $\alpha < 1$ and for every $x > a$ sufficiently close to a the following inequality holds: $|f(x)| \leq (x-a)^{-\alpha}$, then the integral (4) is absolutely convergent; but, if $\alpha \geq 1$ and for every $x > a$ sufficiently close to a we have $f(x) \geq (x-a)^{-\alpha}$, then the integral (4) is divergent.*

The more sensitive test stated by theorem 4 § 107 which enables us sometimes to determine the non-absolute (conditional) convergence of integrals also corresponds to an analogous test for convergence of integrals of unbounded functions which is expressed by the following proposition.

Theorem 4'. *If $\alpha > 0$, the function $\varphi(x)$ is continuous for $x > a$ and there exists a positive number C such that we have for arbitrarily small $\epsilon > 0$*

$$\left| \int_{a+\epsilon}^b \varphi(x) dx \right| < C,$$

then the integral

$$\int_a^b (x-a)^\alpha \varphi(x) dx$$

is convergent.

Proof. Let us assume that

$$\int_u^b \varphi(x) dx = \Phi(u) \quad (a < u \leq b),$$

so that $|\Phi(u)| < C$ ($a < u \leq b$). Integrating by parts we obtain

$$\begin{aligned} \int_{a+\varepsilon}^b (x-a)^\alpha \varphi(x) dx &= \\ &= \left[-(x-a)^\alpha \Phi(x) \right]_{a+\varepsilon}^b + \alpha \int_{a+\varepsilon}^b (x-a)^{\alpha-1} \Phi(x) dx = \\ &= \varepsilon^\alpha \Phi(a+\varepsilon) + \alpha \int_{a+\varepsilon}^b (x-a)^{\alpha-1} \Phi(x) dx, \end{aligned}$$

since $\Phi(b) = 0$. When $\varepsilon \rightarrow 0$, the first term on the right-hand side tends to zero since $\alpha > 0$ and $|\Phi(a+\varepsilon)| < C$. The absolute value of the integrand of the second term is less than $C/(x-a)^{1-\alpha}$, where $1-\alpha < 1$. It therefore follows from theorem 3' that the second term on the right-hand side tends to the following integral as its limit for $\varepsilon \rightarrow 0$:

$$\alpha \int_a^b (x-a)^{\alpha-1} \Phi(x) dx.$$

Hence, if $\varepsilon \rightarrow 0$, the left-hand side of the last equation also has the same limit and theorem 4' is proved.

Example 1. Let us consider the integral

$$I = \int_0^1 \frac{\ln x}{\sqrt{x}} dx.$$

For every constant $\lambda > 0$ the product $x^\lambda \ln x$ tends to zero for $x \rightarrow 0$, as can be seen directly by applying L' Hopital's rule, if we represent this product in the form of the ratio $\ln x/x^{-\lambda}$. Assuming, in particular, that $\lambda = 1/4$, we have for a sufficiently small $x > 0$

$$|\ln x| < x^{-\frac{1}{4}},$$

and consequently

$$\frac{|\ln x|}{\sqrt{x}} < x^{-\frac{3}{4}}$$

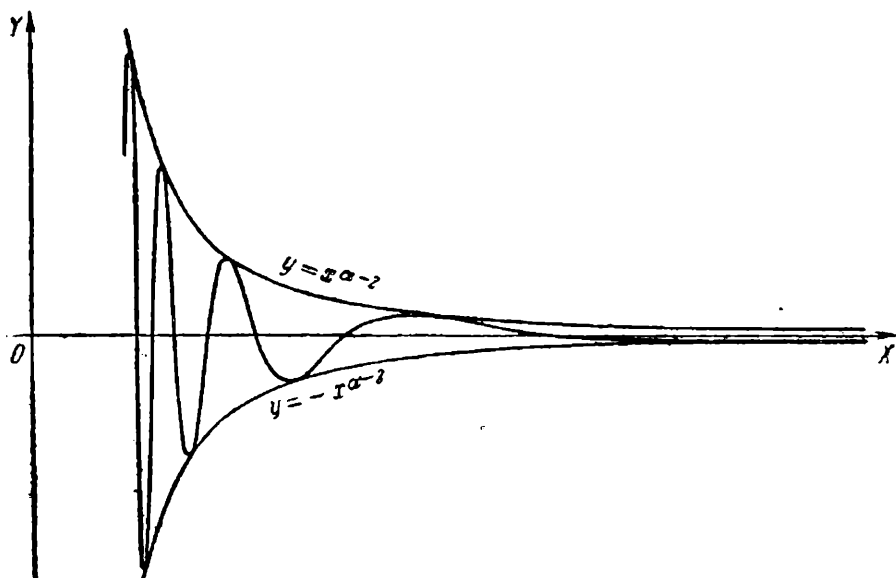


Fig. 68.

and this, according to theorem 3', implies absolute convergence of the integral I .

Example 2. Let us consider the integral

$$I = \int_0^1 \left(\cos \frac{1}{x} \right) \frac{dx}{x^{2-\alpha}},$$

where α is an arbitrary positive number. When α is small, the quantity $1/x^{2-\alpha}$ increases very rapidly as $x \rightarrow 0$; and since $\cos 1/x$ vibrates an infinite number of times between $+1$ and -1 in this process, therefore, the integrand is strictly unbounded in the neighbourhood of the point $x = 0$; the graph of this function is represented in fig. 68.

The replacement of the variable $x = 1/y$ gives:

$$\left| \int_{\varepsilon}^1 \cos \frac{1}{x} \frac{dx}{x^2} \right| = \left| \int_1^{\frac{1}{\varepsilon}} \cos y \, dy \right| = \left| \sin \frac{1}{\varepsilon} - \sin 1 \right| < 2.$$

Therefore the function

$$\varphi(x) = \frac{1}{x^2} \cos \frac{1}{x}$$

satisfies the requirements of theorem 4' in the interval $(0, 1)$. Applying this theorem ($a = 0$) we find that the integral

$$\int_0^1 x^{\alpha} \varphi(x) \, dx = \int_0^1 \cos \frac{1}{x} \frac{dx}{x^{2-\alpha}} = I$$

is convergent for every $\alpha > 0$.

For further exercises cf. Problem Book by B. P. Demidovich, Section IV, Nos. 102, 104, 110.

CHAPTER XXVI

INTEGRALS OF PARAMETRIC FUNCTIONS

§ 109. Integrals with finite limits

The study of quantitative relations existing in the world around us leads to the discovery of many new kinds of functional dependencies. It is one of the main objects of mathematical analysis to investigate more and more classes of functions. However, a mere mention of the types of functions in existence would, of course, be quite insufficient ; we must define such families and develop methods for their study, for otherwise the newly discovered functions would be quite useless : functions and properties of functions which we cannot study are useless to us. Therefore, for the study of new unknown functional dependencies science always tries to provide an instrument which enables us to deal with these functions systematically and develop their main properties gradually.

Our studies so far are rich in examples illustrating these points. The sum of an infinite series of functions whose terms are composed of well-known functions (for example, power and trigonometrical) is, generally speaking, a new function whose properties we hardly know ; but the series which defines it usually appears to be a strong and convenient tool for the study of its main outlines (for example, for its evaluation and definition of the results of operations performed on it). We are already familiar with the useful properties of infinite series and the close connection existing between the properties of the series and functions. Integral calculus provides an even more instructive example. We know that a whole family of primitives exists for every continuous function and even for simple (elementary) functions these primitives are usually new functions about which we know nothing except that they are differentiable and their derivatives are equal to

known functions. It is therefore self-evident that finding the primitives is a useful source for the definition of new functions; however, this definition is by itself almost useless, for although we know about an "integral logarithm" that its derivative is equal to $1 / \ln x$, this tells us nothing about the properties of this function and we also have no method for evaluation of the integral logarithm. How do we overcome this obstacle? We make a superior *apparatus* (integral) which enables us to find primitives of the given functions; we give, as it were, *constructive* definition to primitives, *i.e.* we define the functions by means of an instrument which proves to be very convenient for studying the properties of functions (and which, in principle, enables us to evaluate these functions with any desired degree of accuracy); and historically too, integral calculus only became a convenient method for studying new classes of functional dependencies after this apparatus was created.

In this chapter we shall learn a new method for defining and studying functions; historically it is one of the most productive methods, for it enables us to study in detail many functional dependencies which are of the greatest importance in theory and many diverse applications.

Let the function $f(x, u)$ continuous in the rectangle $R(a \leq x \leq b, \alpha \leq u \leq \beta)$ be a function of two independent variables. Let us choose and fix definite value of the variable u in the interval (α, β) ; $f(x, u)$ thus becomes a continuous function of one variable x in the interval (a, b) ; the integral

$$\int_a^b f(x, u) dx$$

of this function depends in general on the chosen value of the variable u ; it has a definite value for every choice of u and usually changes when u changes; it is therefore a function of u defined in the interval (α, β) ; let us denote it by $\varphi(u)$ so that

$$\varphi(u) = \int_a^b f(x, u) dx \quad (\alpha \leq u \leq \beta). \quad (1)$$

The variable u on which the integrand depends but which is assumed to be constant during integration (having been given a

definite fixed value) is usually called a *parameter*; the value of the integral depends on the chosen value of the parameter; the integrand can, in some cases, evidently depend on a whole series of parameters u_1, u_2, \dots, u_k ; in that case the integral

$$\varphi(u_1, u_2, \dots, u_k) = \int_a^b f(x, u_1, u_2, \dots, u_k) dx \quad (2)$$

is a function of the whole set of these parameters.

Example. We have learnt in § 66 that

$$\int e^{ux} \cos vx dx = \frac{e^{ux} (v \sin vx + u \cos vx)}{u^2 + v^2} + C;$$

we can therefore obtain

$$\int_0^1 e^{ux} \cos vx dx = \frac{e^u (v \sin v + u \cos v) - u}{u^2 + v^2};$$

the integral on the left-hand side of this equation is a function of two parameters u and v ; the right-hand side gives us an elementary expression for this function.

In the above example the integral depending on the parameters u and v is an elementary function of these parameters. However, in most cases an integral of the type (2), even when $f(x, u_1, u_2, \dots, u_k)$ is an elementary function of its constituent variables, is a *non-elementary* function, for whose study we have no other means except the integral (2) itself. Hence we must study the properties of the function φ only with the help of (2); even this function can only be evaluated by means of the integral itself. It is therefore obvious that the properties of these integrals and the rules for analytical operations with them require careful study, to which the rest of this chapter is devoted.

For the sake of simplicity we shall only consider integrals of the type (1) which depend on one parameter only, although our arguments and results can be extended without difficulties to include the general case of integrals of the type (2).

Theorem 1. *If the function $f(x, u)$ is continuous in the rectangle R ($a \leq x \leq b$, $\alpha \leq u \leq \beta$), then the function $\varphi(u)$, defined by the integral (1), is continuous in the interval (α, β) .*

Proof. If the point u and $u + \Delta u$ lie in the interval (α, β) ; then

$$\varphi(u + \Delta u) - \varphi(u) = \int_a^b [f(x, u + \Delta u) - f(x, u)] dx,$$

and therefore

$$|\varphi(u + \Delta u) - \varphi(u)| \leq \int_a^b |f(x, u + \Delta u) - f(x, u)| dx. \quad (3)$$

But it follows from theorem 3 § 88 that the function $f(x, u)$ is uniformly continuous in the rectangle R ; therefore there exists a $\delta > 0$ for arbitrarily small $\varepsilon > 0$ for all points $(x_1, u_1), (x_2, u_2)$ of this rectangle which are at a distance less than δ

$$|f(x_2, u_2) - f(x_1, u_1)| < \varepsilon.$$

Since the distance of the points (x, u) and $(x, u + \Delta u)$ is equal to $|\Delta u|$, therefore, we have for $|\Delta u| < \delta$

$$|f(x, u + \Delta u) - f(x, u)| < \varepsilon,$$

whatever the points $(x, u), (x, u + \Delta u)$ of the rectangle R . Thus for $|\Delta u| < \delta$ on account of (3) we have

$$|\varphi(u + \Delta u) - \varphi(u)| < \varepsilon (b - a) \quad (\alpha \leq u, u + \Delta u \leq \beta);$$

Since $\varepsilon > 0$ can be chosen as small as we please, we have therefore theorem 1 is proved.

Having thus established continuity of the integral (1) with respect to the parameter u we can now integrate the function $\varphi(u)$ in the interval (α, β) (or in any subinterval). The integral

$$\int_{\alpha}^{\beta} \varphi(u) du = \int_{\alpha}^{\beta} \left\{ \int_a^b f(x, u) dx \right\} du \quad (4)$$

always has a definite meaning (provided the function $f(x, u)$ is continuous in the region R). We know that for series of functions (§ 75) “term-by-term” integration is of great importance, i.e. we can integrate under the summation sign; similarly for functions of the type (1) it is most important to integrate with respect to u under the sign of integral with respect to x ; the question therefore arises, whether the integral (4) will coincide with the integral

$$\int_a^b \left\{ \int_{\alpha}^{\beta} f(x, u) du \right\} dx,$$

which differs from it only by the order in which the integrations are performed. Let us note that term-by-term integrability of the series of functions

$$\sum_{k=1}^{\infty} u_k(x)$$

in the interval (a, b) depends on whether the equation

$$\int_a^b \left\{ \sum_{k=1}^{\infty} u_k(x) \right\} dx = \sum_{k=1}^{\infty} \left\{ \int_a^b u_k(x) dx \right\},$$

holds, *i.e.* the interchange of the order of summation with respect to k and integration with respect to x . Similarly the ability to integrate the function $\varphi(u)$ with respect to u under the integral with respect to x is equivalent to the independence of the result of successive integration of the function $f(x, u)$ with respect to x and u from the order in which these operations are carried out. We will now show that this problem can always be solved positively for continuous functions.

Theorem 2. *If the function $f(x, u)$ is continuous in the rectangle R , then*

$$\int_{\alpha}^{\beta} \varphi(u) du = \int_{\alpha}^{\beta} \left\{ \int_a^b f(x, u) dx \right\} du = \int_a^b \left\{ \int_{\alpha}^{\beta} f(x, u) du \right\} dx.$$

Proof. Let $a \leq a' < b' \leq b$, $\alpha \leq \alpha' < \beta' \leq \beta$ and let m and M denote respectively the lower and upper bounds of the function $f(x, u)$ in the rectangle $(a' \leq x \leq b', \alpha' \leq u \leq \beta')$. We therefore have for $\alpha' \leq u \leq \beta'$

$$m(b' - a') \leq \int_{a'}^{b'} f(x, u) dx \leq M(b' - a')$$

and integrating from α' to β'

$$\begin{aligned} m(b' - a')(\beta' - \alpha') &\leq \int_{\alpha'}^{\beta'} \left\{ \int_{a'}^{b'} f(x, u) dx \right\} du \leq \\ &\leq M(b' - a')(\beta' - \alpha'). \end{aligned} \tag{5}$$

Similarly we find that we also have

$$\begin{aligned} m (b' - a') (\beta' - \alpha') &\leq \int_{a'}^{b'} \left\{ \int_{\alpha'}^{\beta'} f(x, u) du \right\} dx \leq \\ &\leq M (b' - a') (\beta' - \alpha'). \end{aligned} \quad (6)$$

Let us now divide the interval (a, b) into n subintervals by means of the following points of division ;

$$a = x_0 < x_1 < x_2 < \dots < x_n = b,$$

and the interval (α, β) into m subintervals by means of the following points of division :

$$\alpha = u_0 < u_1 < u_2 < \dots < u_m = \beta.$$

Let us denote by $\Delta_{ik} = (x_i - x_{i-1})(u_k - u_{k-1})$ the area of the rectangle $x_{i-1} \leq x \leq x_i, u_{k-1} \leq u \leq u_k$ and by m_{ik} and M_{ik} respectively the upper and lower bounds of the function $f(x, u)$ in this rectangle. Applying the inequalities (5) and (6) to this rectangle we obtain

$$m_{ik} \Delta_{ik} \leq \int_{u_{k-1}}^{u_k} \left\{ \int_{x_{i-1}}^{x_i} f(x, u) dx \right\} du \leq M_{ik} \Delta_{ik}, \quad (7)$$

$$\begin{aligned} m_{ik} \Delta_{ik} &\leq \int_{x_{i-1}}^{x_i} \left\{ \int_{u_{k-1}}^{u_k} f(x, u) du \right\} dx \leq M_{ik} \Delta_{ik} \\ (1 \leq i \leq n, 1 \leq k \leq m) \end{aligned} \quad (8)$$

We now note that

$$\begin{aligned} I &= \int_{\alpha}^{\beta} \left\{ \int_a^b f(x, u) dx \right\} du = \sum_{k=1}^m \int_{u_{k-1}}^{u_k} \left\{ \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x, u) dx \right\} du = \\ &= \sum_{i=1}^n \sum_{k=1}^m \int_{u_{k-1}}^{u_k} \left\{ \int_{x_{i-1}}^{x_i} f(x, u) dx \right\} du, \end{aligned}$$

and similarly

$$I' = \int_a^b \left\{ \int_{\alpha}^{\beta} f(x, u) du \right\} dx = \sum_{i=1}^n \sum_{k=1}^m \int_{x_{i-1}}^{x_i} \left\{ \int_{u_{k-1}}^{u_k} f(x, u) du \right\} dx.$$

Applying the inequalities (7) and (8) respectively to all terms of the double sums of these equations we obtain:

$$\sum_{i=1}^n \sum_{k=1}^m m_{ik} \Delta_{ik} \leq I \leq \sum_{i=1}^n \sum_{k=1}^m M_{ik} \Delta_{ik},$$

$$\sum_{i=1}^n \sum_{k=1}^m m_{ik} \Delta_{ik} \leq I' \leq \sum_{i=1}^n \sum_{k=1}^m M_{ik} \Delta_{ik}.$$

Hence the integrals I and I' whose equality we must prove are both confined between these two limits; the absolute value of the difference $I - I'$ therefore cannot exceed the distance between these two limits, *i.e.*

$$|I - I'| \leq \sum_{i=1}^n \sum_{k=1}^m (M_{ik} - m_{ik}) \Delta_{ik}.$$

If the above divisions of the intervals (a, b) and (α, β) are sufficiently fine (and there is no reason why we should not make them as fine as possible), then uniform continuity of the function $f(x, u)$ in the rectangle R implies that all the differences $M_{ik} - m_{ik}$ ($1 \leq i \leq n$, $1 \leq k \leq m$) will be less than an arbitrary preassigned positive number ε . Therefore

$$|I - I'| \leq \varepsilon \sum_{i=1}^n \sum_{k=1}^m \Delta_{ik} = \varepsilon (b - a)(\beta - \alpha).$$

Since $\varepsilon > 0$ is as small as we please and the left-hand side is independent of ε , therefore $I = I'$ and theorem 2 is proved.

We shall now consider differentiability of the function $\varphi(u)$ given by the integral (1). Just as in term-by-term differentiation of infinite series we are here concerned with differentiability of the function $\varphi(u)$ in the interval (α, β) and possibility of expressing its derivative in the form

$$\int_a^b \frac{\partial f(x, u)}{\partial u} dx,$$

or, as it is usually said, with “differentiation under the sign of the integral”. If differentiability of the terms of the series was a neces-

sary condition, so now we must also assume existence and continuity (or at least integrability with respect to x) of the the partial derivative $\partial f / \partial u$ in the rectangle R : However, this assumption is also sufficient as can be seen from the following theorem.

Theorem 3. *If the function $f(x, u)$ and its partial derivative $\partial f(x, u) / \partial u$ are continuous in the rectangle R , then the function $\varphi(u)$, given by the interval (1), is differentiable in the interval (α, β) and*

$$\varphi'(u) = \int_a^b \frac{\partial f(x, u)}{\partial u} dx \quad (\alpha \leq u \leq \beta). \quad (9)$$

Proof. Let us assume that

$$\int_a^b \frac{\partial f(x, u)}{\partial u} dx = g(u) \quad (\alpha \leq u \leq \beta).$$

It follows from theorem 2 that for $\alpha \leq v \leq \beta$

$$\begin{aligned} \int_{\alpha}^v g(u) du &= \int_a^b \left\{ \int_{\alpha}^v \frac{\partial f(x, u)}{\partial u} du \right\} dx = \\ &= \int_a^b \{f(x, v) - f(x, \alpha)\} dx = \varphi(v) - \varphi(\alpha). \end{aligned}$$

The left-hand side of this inequality which is an integral of a continuous function is integrable with respect to the upper limit v and its derivative is equal to $g(v)$ (theorem 1 § 50); therefore $\varphi'(v)$ exists and is equal to $g(v)$ for every $v(\alpha \leq v \leq \beta)$. Theorem 3 is thus proved.

We have assumed in all that is said above that the limits a and b of the integral (1) are constant. However, it often happens in practice that we integrate within different limits for different values of the parameter u so that a and b become functions of the parameter u : $a = a(u)$, $b = b(u)$. Evidently such an integral

$$\psi(u) = \int_{a(u)}^{b(u)} f(x, u) dx, \quad (10)$$

like the integral (1), is a function of the parameter u .

We shall now consider some properties of this more general dependency of the integral on the parameter. We shall always assume in such cases that the function $f(x, u)$ is continuous in the rectangle R and the functions $a(u)$ and $b(u)$ are continuous in the interval (α, β) where

$$a \leq a(u) \leq b, \quad a \leq b(u) \leq b \quad (\alpha \leq u \leq \beta).$$

It can be shown that the function $\psi(u)$ is continuous in the interval (α, β) . In fact, if u and $u + \Delta u$ lie in this interval, then

$$\psi(u + \Delta u) - \psi(u) = \int_{a(u + \Delta u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx - \int_{a(u)}^{b(u)} f(x, u) dx.$$

Since

$$\begin{aligned} & \int_{a(u + \Delta u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx = \\ & = \int_{a(u + \Delta u)}^{a(u)} f(x, u + \Delta u) dx + \int_{a(u)}^{b(u)} f(x, u + \Delta u) dx + \\ & \quad + \int_{b(u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx, \end{aligned}$$

therefore

$$\begin{aligned} \psi(u + \Delta u) - \psi(u) &= \int_{a(u + \Delta u)}^{a(u)} f(x, u + \Delta u) dx + \\ &+ \int_{b(u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx + \int_{a(u)}^{b(u)} [f(x, u + \Delta u) - f(x, u)] dx. \quad (11) \end{aligned}$$

When $\Delta u \rightarrow 0$, the limits of the last integral on the right-hand side remain constant; the argument used for deducing theorem 1 therefore shows that it tends to zero. The absolute values of the first two integrals are evidently less than

$$M |a(u + \Delta u) - a(u)|, \quad M |b(u + \Delta u) - b(u)|$$

(where M is the upper bound of the function $|f(x, y)|$ in the rectangle R) and it therefore follows from the assumed continuity that the functions $a(u)$ and $b(u)$ also tend to zero as $\Delta u \rightarrow 0$.

Hence

$$\psi(u + \Delta u) - \psi(u) \rightarrow 0 \quad (\Delta u \rightarrow 0) \quad (\alpha \leq u \leq \beta),$$

and thus the function $\psi(u)$ is continuous in the interval (α, β) ; the following generalisation of theorem 1 therefore holds:

Theorem 4. *If the function $f(x, u)$ is continuous in the rectangle R and the functions $a(u)$ and $b(u)$ are continuous in the interval (α, β) , where*

$$a \leq a(u) \leq b, \quad a \leq b(u) \leq b \quad (\alpha \leq u \leq \beta),$$

then the function $\psi(u)$, given by the integral (10), is continuous in the interval (α, β) .

We will now assume that the functions $a(u)$ and $b(u)$ are not only continuous but also differentiable in the interval (α, β) and the function $f(x, u)$ has a continuous partial derivative $\partial f / \partial u$ in the rectangle R ; we will show that the function $\psi(u)$ is in this case differentiable in the interval (a, b) and $\psi'(u)$ can be expressed by a simple formula which is the direct generalisation of formula (9).

We have:

$$\begin{aligned} \psi(u + \Delta u) - \psi(u) &= \int_{a(u + \Delta u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx - \int_{a(u)}^{b(u)} f(x, u) dx = \\ &= \left\{ \int_{a(u)}^{b(u)} f(x, u + \Delta u) dx - \int_{a(u)}^{b(u)} f(x, u) dx \right\} + \\ &+ \int_{b(u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx - \int_{a(u)}^{a(u + \Delta u)} f(x, u + \Delta u) dx. \end{aligned} \quad (12)$$

We will consider u as constant and assume that

$$\varphi(v) = \int_{a(u)}^{b(u)} f(x, v) dx \quad (\alpha \leq v \leq \beta).$$

According to theorem 3 the function $\varphi(v)$ is differentiable in the interval (α, β) and

$$\varphi'(v) = \int_{a(u)}^{b(u)} \frac{\partial f(x, v)}{\partial v} dx \quad (\alpha \leq v \leq \beta);$$

in particular, when $v = u$,

$$\begin{aligned} \int_{a(u)}^{b(u)} \frac{\partial f(x, u)}{\partial u} dx &= \varphi'(u) = \lim_{\Delta u \rightarrow 0} \frac{\varphi(u + \Delta u) - \varphi(u)}{\Delta u} = \\ &= \lim_{\Delta u \rightarrow 0} \frac{1}{\Delta u} \left\{ \int_{a(u)}^{b(u)} f(x, u + \Delta u) dx - \int_{a(u)}^{b(u)} f(x, u) dx \right\}. \end{aligned} \quad (13)$$

This shows that the first term on the right-hand side of the equations (12) if divided by Δu tends to the following integral as its limit for $\Delta u \rightarrow 0$:

$$\int_{a(u)}^{b(u)} \frac{\partial f(x, u)}{\partial u} dx.$$

Let us now consider the second term. According to the mean-value theorem

$$\int_{b(u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx = f(\xi, u + \Delta u) [b(u + \Delta u) - b(u)],$$

where ξ is confined between $b(u)$ and $b(u + \Delta u)$. Therefore

$$\frac{1}{\Delta u} \int_{b(u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx = f(\xi, u + \Delta u) \frac{b(u + \Delta u) - b(u)}{\Delta u}.$$

When $\Delta u \rightarrow 0$, the second factor on the right-hand side tends to $b'(u)$. However, in the first factor $u + \Delta u \rightarrow u$, $\xi \rightarrow b(u)$ and it therefore follows from continuity of the function $f(x, u)$ that

$$f(\xi, u + \Delta u) \rightarrow f[b(u), u].$$

Therefore

$$\lim_{\Delta u \rightarrow 0} \frac{1}{\Delta u} \int_{b(u)}^{b(u + \Delta u)} f(x, u + \Delta u) dx = f[b(u), u] b'(u). \quad (14)$$

We have similarly

$$\lim_{\Delta u \rightarrow 0} \frac{1}{\Delta u} \int_{a(u)}^{a(u + \Delta u)} f(x, u + \Delta u) dx = f[a(u), u] a'(u). \quad (15)$$

Finally taking into consideration the results (13), (14), (15) we obtain from the equation (12) :

$$\begin{aligned} \lim_{\Delta u \rightarrow 0} \frac{\psi(u + \Delta u) - \psi(u)}{\Delta u} &= \\ &= \int_{a(u)}^{b(u)} \frac{\partial f(x, u)}{\partial u} dx + f[b(u), u] b'(u) - f[a(u), u] a'(u). \end{aligned}$$

We have thus proved differentiability of the function $\psi(u)$ at the point u and found an expression for the derivative $\psi'(u)$. The result can be expressed by the following proposition.

Theorem 5. *If the function $f(x, u)$ is continuous in the rectangle R and if it has a continuous partial derivative with respect to u and the functions $a(u)$ and $b(u)$ are differentiable in the interval (α, β) and*

$$a \leq a(u) \leq b, \quad a \leq b(u) \leq b \quad (\alpha \leq u \leq \beta),$$

then the function $\psi(u)$, given by the integral (10), is differentiable in the interval (α, β) and

$$\psi'(u) = \int_{a(u)}^{b(u)} \frac{\partial f(x, u)}{\partial u} dx + f[b(u), u] b'(u) - f[a(u), u] a'(u). \quad (16)$$

In particular, if $a(u)$ and $b(u)$ are constant in the interval (α, β) , we have $a'(u) = b'(u) = 0$ and formula (16) becomes formula (9) of theorem 3.

For exercises, cf. Problem Book by B.P. Demidovich, Section VII, Nos. 17, 18, 23.

§ 110. Integrals with infinite limits

In the last paragraph we have considered integrals with finite limits which depended on parameters and shown the analogy between problems and arguments in this field and corresponding problems in the theory of infinite series. However, this analogy only becomes fully apparent by studying integrals with infinite limits which depend on parameters, *i.e.* by studying integrals of the type

$$\int_a^{\infty} f(x, u) dx \quad (1)$$

(as in chapter 25, we shall restrict ourselves to the consideration of integrals with infinite *upper* limits; it is needless to say that all properties of these integrals can be symmetrically extended to integrals with infinite *lower* limits and, subsequently, to integrals with *both* limits infinite). As we shall see, the theory of these integrals serves as the basis for the deduction of many important classical analytical formulae which we shall study in this paragraph.

Let us assume that the integral (1) is convergent for all values of the parameter u in the interval $\alpha \leq u \leq \beta$. It is therefore a function of the parameter u defined in the interval (α, β) :

$$\int_a^{\infty} f(x, u) dx = \varphi(u). \quad (2)$$

We know from the theory of series of functions (chapter 19) that the concept to *uniform convergence* is of fundamental importance for these series; for integrals of the type (1) the analogous concept is equally important. If the integral (1) is convergent at any arbitrary point u in the interval (α, β) , then for every $\varepsilon > 0$ and for every point u in (α, β) a number A_0 can be found (which depends on ε and u) such that for every $A \geq A_0$

$$\left| \int_A^{\infty} f(x, u) dx \right| < \varepsilon. \quad (3)$$

For a given ε the number A_0 will, in general, be different for different values of u . If for every $\varepsilon > 0$ there exists an A_0 such that for $A \geq A_0$ the inequality (3) holds for every u ($\alpha \leq u \leq \beta$), we say that the integral (1) is *uniformly convergent* in (α, β) .

Example. Let us consider the integral

$$I(\alpha) = \int_0^{\infty} e^{-\alpha x} \frac{\sin x}{x} dx,$$

where $\alpha \geq 0$. We have seen at the end of § 66 that

$$\int e^{ax} \sin bx dx = \frac{e^{ax} (a \sin bx - b \cos bx)}{a^2 + b^2} + C,$$

where a and b are arbitrary constants. In particular,

$$\int e^{-\alpha x} \sin x dx = \frac{-e^{-\alpha x} (\alpha \sin x + \cos x)}{1 + \alpha^2} + C = \Phi_{\alpha}(x) + C,$$

where the function $\Phi_{\alpha}(x)$ evidently remains bounded for $x \geq 0$, $\alpha \geq 0$:

$$|\Phi_{\alpha}(x)| < B \quad (x \geq 0, \alpha \geq 0),$$

where B is a constant. Integrating by parts we therefore obtain for $\alpha > 0$: *)

$$\begin{aligned} \left| \int_a^{\infty} e^{-\alpha x} \frac{\sin x}{x} dx \right| &= \left| \left[\frac{\Phi_{\alpha}(x)}{x} \right]_a^{\infty} + \int_a^{\infty} \frac{\Phi_{\alpha}(x)}{x^2} dx \right| \leq \\ &\leq \left| \frac{\Phi_{\alpha}(a)}{a} \right| + \int_a^{\infty} \frac{|\Phi_{\alpha}(x)|}{x^2} dx < \frac{2B}{a}. \end{aligned}$$

Since the right-hand side is independent of α and tends to zero as $a \rightarrow \infty$, therefore the integral $I(\alpha)$ is uniformly convergent with respect to the parameter α in the semi-straight line $\alpha \geq 0$.

For this class of integrals the concept of uniform convergence is just as important as for series of functions. We shall soon confirm this by a series of propositions each of which corresponds to an analogous theorem in chapter 19.

*) Formulae for integration by parts can be applied to integrals with infinite limits only when both constituent integrals are convergent (in such cases the term free of integrals must tend to a definite limit when the independent variable increases indefinitely). This can be shown directly by writing the formula for a finite interval of integration (a, b) and assuming subsequently that b increases indefinitely.

We shall at first consider a simple and convenient test for convergence which is analogous to theorem 2 § 73:

Test for uniform convergence. *If a continuous function $F(x)$ exists such that $|f(x, u)| \leq F(x)$ for all sufficiently large values of u in the interval (α, β) and if the integral*

$$\int_a^\infty F(x) dx$$

is convergent, then the integral (2) is uniformly convergent in the interval (α, β) .

The proof is analogous to that of theorem 2 § 73 and based on the inequalities

$$\left| \int_A^B f(x, u) dx \right| \leq \int_A^B |f(x, u)| dx \leq \int_A^B F(x) dx,$$

which hold for a sufficiently large A and $B > A$.

In the same way as uniform convergence of a series of continuous functions guarantees continuity of its sum (theorem 1 § 74), so uniform convergence of the integral (2) (when the function $f(x, u)$ is continuous) implies continuity of the function of the parameter u which it expresses.

Theorem 1. *If the function $f(x, u)$ is continuous for $a \leq x$, $\alpha \leq u \leq \beta$, and the integral (2) is uniformly convergent in the interval (α, β) , then the function $\varphi(u)$ is continuous in that interval.*

The proof is analogous to that of theorem 1 § 74. Let $\varepsilon > 0$, be as small as we please; let us choose A_0 so large that $A_0 > a$ and

$$\left| \int_{A_0}^\infty f(x, u) dx \right| < \frac{\varepsilon}{3} \quad (\alpha \leq u \leq \beta). \quad (4)$$

Since, according to theorem 1 § 109, the integral

$$\int_a^{A_0} f(x, u) dx$$

is a continuous function of u in the interval (α, β) therefore for a sufficiently small $|\Delta u|$

$$\left| \int_a^{A_0} f(x, u + \Delta u) dx - \int_a^{A_0} f(x, u) dx \right| < \frac{\varepsilon}{3}. \quad (5)$$

But

$$\begin{aligned} \varphi(u + \Delta u) - \varphi(u) = & \left\{ \int_a^{A_0} f(x, u + \Delta u) dx - \int_a^{A_0} f(x, u) dx \right\} + \\ & + \int_{A_0}^{\infty} f(x, u + \Delta u) dx - \int_{A_0}^{\infty} f(x, u) dx, \end{aligned}$$

it therefore follows from (4) and (5) that when $|\Delta u|$ is sufficiently small

$$\begin{aligned} |\varphi(u + \Delta u) - \varphi(u)| \leq & \left| \int_a^{A_0} f(x, u + \Delta u) dx - \int_a^{A_0} f(x, u) dx \right| + \\ & + \left| \int_{A_0}^{\infty} f(x, u + \Delta u) dx \right| + \left| \int_{A_0}^{\infty} f(x, u) dx \right| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

Theorem 1 is thus proved.

Uniform convergence of the integral (2) which is a sufficient condition for continuity of the function $\varphi(u)$ is not, however, the necessary condition for this purpose; the position is just as with series. There is one important particular case, however, when it is also the necessary condition; we have the following proposition which is analogous to theorem 2 § 74.

Theorem 2. *If $f(x, u)$ is continuous and has a constant sign in the region $x \geq a$, $\alpha \leq u \leq \beta$, then continuity of the function $\varphi(u)$ in the interval (α, β) implies uniform convergence of the integral (2).*

Proof. Let us assume, say, that $f(x, u) \geq 0$ ($x \geq a$, $\alpha \leq u \leq \beta$). Let u_0 be an arbitrary point in the interval (α, β) . Since the integral (2) is convergent at the point u_0 , therefore for arbitrarily small $\varepsilon > 0$ a number A_0 exists such that

$$\int_{A_0}^{\infty} f(x, u_0) dx < \varepsilon; \quad (6)$$

where the number A_0 depends on the chosen point u_0 ; but since, according to our assumptions, the function $\varphi(u)$ is continuous in the interval (α, β) and it also follows from theorem 1 § 109 that the func-

tion $\int_a^{A_0} f(x, u) dx$ is continuous in that same interval, therefore the function

$$\int_{A_0}^{\infty} f(x, u) dx = \varphi(u) - \int_a^{A_0} f(x, u) dx$$

is also continuous in (α, β) ; hence the inequality (6) which is satisfied at the point u_0 must also be satisfied in a neighbourhood of that point. Therefore every point in the interval (α, β) also lies in an interval all points of which satisfy the inequality (6). The set of all subintervals constructed for all the points u_0 in the intervals (α, β) covers entirely the interval (α, β) . In accordance with the lemma on finite coverage a finite group $\delta_1, \delta_2, \dots, \delta_n$ of subintervals of this system exists which also covers the entire interval (α, β) . Each subinterval δ_i ($1 \leq i \leq n$) corresponds to a number A_i so that

$$\int_{A_i}^{\infty} f(x, u) dx < \varepsilon$$

for all the points u in the subinterval δ_i . Let A be the greatest of the numbers A_1, A_2, \dots, A_n ; it then follows from the condition $f(x, u) \geq 0$ ($x \geq a, \alpha \leq u \leq \beta$) that:

$$\int_A^{\infty} f(x, u) dx < \varepsilon$$

for all the points u in any one of the subintervals δ_i and, consequently, also for all points u in the interval (α, β) since this interval is covered by the system of subintervals $\delta_1, \delta_2, \dots, \delta_n$. And since $\varepsilon > 0$ is arbitrarily small therefore the integral

$$\int_a^\infty f(x, u) dx$$

is uniformly convergent in the interval (α, β) . Theorem 2 is thus proved.

Moreover, just as for the term-by-term integration of uniformly convergent series of functions the following proposition holds :

Theorem 3. *Under the conditions stated in theorem 1 we have*

$$\int_\alpha^\beta \varphi(u) du = \int_a^\infty \left\{ \int_\alpha^\beta f(x, u) du \right\} dx.$$

In other words, under the given conditions an interchange of integration with respect to x and u is permissible :

$$\int_\alpha^\beta \left\{ \int_a^\infty f(x, u) dx \right\} du = \int_a^\infty \left\{ \int_\alpha^\beta f(x, u) du \right\} dx,$$

i.e. in the case of uniform convergence theorem 2 § 109 can be extended to include the case $b = +\infty$.

Proof. Let the number $A_0 > a$ be so great that for $A \geq A_0$

$$\left| \int_A^\infty f(x, u) dx \right| < \varepsilon \quad (\alpha \leq u \leq \beta). \quad (7)$$

We have :

$$\int_\alpha^\beta \varphi(u) du = \int_\alpha^\beta \left\{ \int_a^A f(x, u) dx \right\} du + \int_\alpha^\beta \left\{ \int_A^\infty f(x, u) dx \right\} du$$

But it follows from theorem 2 § 109

$$\int_\alpha^\beta \left\{ \int_a^A f(x, u) dx \right\} du = \int_a^A \left\{ \int_\alpha^\beta f(x, u) du \right\} dx,$$

and as a result of the evaluation (7) the above equation gives :

$$\left| \int_\alpha^\beta \varphi(u) du - \int_a^A \left\{ \int_\alpha^\beta f(x, u) du \right\} dx \right| < \varepsilon (\beta - \alpha)$$

provided only $A \geq A_0$; but this means that the integral

$$\int_a^\infty \left\{ \int_\alpha^\beta f(x, u) du \right\} dx$$

is convergent and equal to $\int_\alpha^\beta \varphi(u) du$. Theorem 3 is thus proved.

In the above two integrations whose order can, in accordance with theorem 3, be reversed, we have only one infinite limit whereas the other limit is finite. In practice the following much more complicated case is of greater interest, *i.e.* when the limits of both integrals are infinite. A typical example of this type is held in the question, under which conditions we have

$$\int_a^\infty \left\{ \int_\alpha^\infty f(x, u) du \right\} dx = \int_\alpha^\infty \left\{ \int_a^\infty f(x, u) dx \right\} du. \quad (8)$$

In this book we cannot consider this problem thoroughly. We shall restrict ourselves to the investigation of a particular case which will be found useful later and whose conditions are suitable for deducing tests which would make the transposition (8) possible; we shall assume that the function $f(x, u)$ has a constant sign in the whole region $x \geq a$, $u \geq \alpha$; we shall assume, say, that it is non-negative. In that case the following theorem holds:

Theorem 4. *If the function $f(x, u) \geq 0$ is continuous in the region $x \geq a$, $u \geq \alpha$ and if the functions ^{*}*

$$\varphi(u) = \int_a^\infty f dx, \quad \psi(x) = \int_\alpha^\infty f du \quad (9)$$

are respectively continuous for $u \geq \alpha$ and $x \geq a$ and out of the integrals

$$\int_a^\infty \psi(x) dx, \quad \int_\alpha^\infty \varphi(u) du$$

*at least one is convergent, then the other is also convergent and the two integrals are equal to one another *i.e.* the equation (8) holds.*

^{*}) Here and in future we shall write f instead of $f(x, y)$ to simplify notation.

Proof. Let us note that as a result of theorem 2, the assumed continuity of the functions $\varphi(u)$ and $\psi(x)$ implies uniform convergence of both the integrals (9)—the first is uniformly convergent in the arbitrary finite interval $\alpha \leq u \leq \beta$ and the second in the arbitrary interval $a \leq x \leq b$. Let us assume that the following integral is convergent

$$\int_a^\infty \psi(x) dx = I;$$

we must therefore prove that the integral

$$\int_\alpha^\beta \varphi(u) du$$

tends to I as $\beta \rightarrow \infty$, or, as a result of theorem 3

$$\int_\alpha^\beta \varphi(u) du = \int_\alpha^\beta \left\{ \int_a^\infty f dx \right\} du = \int_a^\infty \left\{ \int_\alpha^\beta f du \right\} dx,$$

so that

$$\lim_{\beta \rightarrow \infty} \int_a^\infty \left\{ \int_\alpha^\beta f du \right\} dx = I. \quad (10)$$

Let ε be an arbitrary positive number. The assumed convergence of the integral

$$\int_a^\infty \psi(x) dx$$

implies existence of a number $b > a$ so that

$$\int_b^\infty \psi(x) dx = \int_b^\infty \left\{ \int_\alpha^\infty f du \right\} dx < \varepsilon,$$

and, therefore, as a result of $f(x, u) \geq 0$ we also have for every $\beta > \alpha$

$$\int_b^\infty \left\{ \int_\beta^\infty f du \right\} dx < \varepsilon; \quad (11)$$

Let us now choose the number β so large that

$$\int_{\beta}^{\infty} f \, du < \frac{\varepsilon}{b-a} \quad (a \leq x \leq b)$$

(we can do this, since the second of the integrals (9) is convergent in the interval $(a \leq x \leq b)$). In this case

$$\int_a^b \left\{ \int_{\beta}^{\infty} f \, du \right\} dx < \varepsilon. \quad (12)$$

Adding the inequalities (11) and (12) we have:

$$\int_a^{\infty} \left\{ \int_{\beta}^{\infty} f \, du \right\} dx < 2\varepsilon,$$

or, which is the same

$$I - \int_a^{\infty} \left\{ \int_{\alpha}^{\beta} f \, du \right\} dx < 2\varepsilon,$$

provided β is sufficiently large. Since ε is arbitrary, this proves the relation (10) and therefore also theorem 4.

Let us finally consider differentiability of the function $\varphi(u)$, given by the integral (2), under the sign of the integral, i.e. under what conditions is the function $\varphi(u)$ differentiable and

$$\varphi'(u) = \int_a^{\infty} \frac{\partial f(x, u)}{\partial u} \, dx \quad (\alpha \leq u \leq \beta). \quad (13)$$

The following proposition holds here, which is analogous to the corresponding theorem on term-by-term differentiation of series of functions (theorem 3 § 75):

Theorem. 5. *If in the region $x \geq a$, $\alpha \leq u \leq \beta$ the function $f(x, u)$ is continuous and has a continuous partial derivative $\partial f(x, u) / \partial u$ and if the integral*

$$\int_a^{\infty} \frac{\partial f(x, u)}{\partial u} \, dx \quad (14)$$

is uniformly convergent in the interval (α, β) , then the function $\varphi(u)$, given by the integral (2), is differentiable in that interval and the relation (13) holds.

Let us assume that for every natural number $n > a$

$$\varphi_n(u) = \int_a^n f(x, u) dx,$$

so that we have for $n \rightarrow \infty$:

$$\varphi_n(u) \rightarrow \varphi(u) \quad (\alpha \leq u \leq \beta).$$

According to theorem 3 § 109 the function $\varphi_n(u)$ is differentiable in the interval (α, β) and

$$\varphi'_n(u) = \int_a^n \frac{\partial f(x, u)}{\partial u} dx;$$

uniform convergence of the integral (14) in the interval (α, β) therefore shows that we have uniformly in that interval

$$\varphi'_n(u) \rightarrow \int_a^\infty \frac{\partial f(x, u)}{\partial u} dx \quad (n \rightarrow \infty).$$

It follows from theorem 3 § 75 that the derivative $\varphi'(u)$ exists in the interval (α, β) and coincides with the integral (14), which was to be shown.

Note. Out of all generalised integrals which depend on parameters we have only considered integrals with infinite limits. However, all that we have proved above also holds and is proved by exactly the same methods for the second type of generalised integrals, *i.e.* for integrals of unbounded functions; within the scope of this book we are not even able to state the corresponding theorems; however, there is hardly any need since the analogy is so close that every argument can almost automatically be extended from one case to the other. The reader should find it very useful to define by himself all concepts and statements and give detailed proofs of all theorems for integrals of unbounded functions given in this paragraph.

§ 111. Examples

In §§ 109 and 110 we have dealt with the theory of integrals which depend on parameters and find many applications in analysis.

In the next two paragraphs we shall consider several examples of these applications. It frequently happens that we are given an integral which does not contain parameters but in order to evaluate it is usually simplest to consider this integral in conjunction with other integrals which depend on parameters and use the properties of the latter to evaluate the former. We shall consider several examples of this kind in this paragraph.

Example 1. Evaluate

$$1 = \int_0^{\infty} \frac{1 - e^{-t}}{t} \cos t \, dt.$$

Let us consider the more general integral

$$I(\lambda) = \int_0^{\infty} \frac{1 - e^{-\lambda t}}{t} \cos t \, dt,$$

where λ is an arbitrary non-negative number so that

$$I = I(1);$$

we shall show that the integral $I(\lambda)$ is uniformly convergent with respect to λ in the interval $0 \leq \lambda \leq 1$. The integration in any finite interval $(0, a)$, for example in the interval $(0, 1)$ causes no difficulties since as a result of $0 \leq 1 - e^{-\lambda t} \leq \lambda t$, the integrand remains uniformly bounded for $t \rightarrow 0$ if $0 \leq \lambda \leq 1$.

We must therefore prove that the integral

$$\int_1^{\infty} \frac{1 - e^{-\lambda t}}{t} \cos t \, dt$$

is uniformly convergent with respect to λ in the interval $0 \leq \lambda \leq 1$. But this follows directly from the fact that 1) the integral

$$\int_1^{\infty} \frac{\cos t}{t} \, dt$$

is convergent (§ 107) and is independent of λ and 2) the integral

$$\int_1^{\infty} e^{-\lambda t} \frac{\cos t}{t} \, dt$$

is uniformly convergent on the semi-axis $\lambda \geq 0$; the latter is similarly proved for the example of the integral considered in § 110

$$\int_0^{\infty} e^{-\lambda t} \frac{\sin t}{t} dt \quad (\lambda \geq 0).$$

Hence the integral $I(\lambda)$ is uniformly convergent in the interval $0 \leq \lambda \leq 1$ and is therefore a continuous function of λ in that interval. In particular, when $\lambda \rightarrow 0$

$$I(\lambda) \rightarrow I(0) = 0,$$

we shall find this useful later.

Let us denote by $f(t, \lambda)$ the integrand of the integral $I(\lambda)$. Since

$$\frac{\partial f(t, \lambda)}{\partial \lambda} = e^{-\lambda t} \cos t,$$

therefore the integral

$$\int_0^{\infty} \frac{\partial f(t, \lambda)}{\partial \lambda} dt$$

is uniformly convergent in the interval $\lambda \in (2, 2\lambda)$ for every $\lambda > 0$; therefore $I'(\lambda)$ exists for every $\lambda > 0$ and

$$I'(\lambda) = \int_0^{\infty} e^{-\lambda t} \cos t dt \quad (\lambda > 0)$$

as a result of theorem 5 § 110. But we know that the function $e^{-\lambda t} \cos t$ (§ 66) has a primitive

$$\frac{e^{-\lambda t} (\sin t - \lambda \cos t)}{1 + \lambda^2},$$

which tends to zero as $t \rightarrow \infty$. Therefore for every $\lambda > 0$

$$I'(\lambda) = \left. \frac{e^{-\lambda t} (\sin t - \lambda \cos t)}{1 + \lambda^2} \right|_0^{\infty} = \frac{\lambda}{1 + \lambda^2}$$

and, consequently, since $I(0) = 0$

$$I(\lambda) = I(\lambda) - I(0) = \int_0^{\lambda} I'(x) dx = \int_0^{\lambda} \frac{x}{1 + x^2} dx = \frac{1}{2} \ln(1 + \lambda^2),$$

hence also

$$I = I(1) = \frac{1}{2} \ln 2.$$

Example 2. In § 107 we have proved convergence of the integral

$$I = \int_0^{\infty} \frac{\sin x}{x} dx;$$

we shall now try to evaluate it.

The integral considered in § 110

$$I(\alpha) = \int_0^{\infty} e^{-\alpha x} \frac{\sin x}{x} dx \quad (\alpha \geq 0)$$

becomes I when $\alpha = 0$; we have seen in § 110 that the integral $I(\alpha)$ is uniformly convergent on the whole semi-straight line $\alpha \geq 0$; it follows from theorem 1 § 110 that the function $I(\alpha)$ is therefore continuous for $\alpha \geq 0$; in particular,

$$I = I(0) = \lim_{\alpha \rightarrow +0} I(\alpha).$$

On the other hand, differentiation of the integral $I(\alpha)$ with respect to α under the sign of the integral gives us the integral

$$- \int_0^{\infty} e^{-\alpha x} \sin x dx,$$

which, as can readily be seen, is uniformly convergent for $\alpha \geq \varepsilon$ and for arbitrarily small $\varepsilon > 0$ as a result of the test in § 110 (since $|e^{-\alpha x} \sin x| \leq e^{-\varepsilon x}$). It therefore follows from theorem 5 § 110 that $I'(\alpha)$ exists for $\alpha > 0$ and

$$I'(\alpha) = - \int_0^{\infty} e^{-\alpha x} \sin x dx = - \frac{1}{1 + \alpha^2}$$

(the last equation is established directly, since the primitive $\Phi_{\alpha}(x)$ is known (cf. § 110). Taking into account that $I(\alpha) \rightarrow 0$ as $\alpha \rightarrow \infty$ this gives on integration: *)

*) Since $\left| \frac{\sin x}{x} \right| \leq 1$, therefore $|I(\alpha)| \leq \int_0^{\infty} e^{-\alpha x} dx = \frac{1}{\alpha}$.

$$-I(\alpha) = -\int_{\alpha}^{\infty} \frac{du}{1+u^2} = -\frac{\pi}{2} + \arctan \alpha,$$

and therefore in the limit for $\alpha \rightarrow 0$

$$I(0) = I = \frac{\pi}{2};$$

hence

$$\int_0^{\infty} \frac{\sin x}{x} dx = \frac{\pi}{2},$$

and our problem is solved.

Moreover, noting that integration by parts gives for $a > 0$

$$\int_0^a \frac{\sin x}{x} dx = \left[\frac{1 - \cos x}{x} \right]_0^a + \int_0^a \frac{1 - \cos x}{x^2} dx$$

and that

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x} = 0,$$

we obtain in the limit for $a \rightarrow \infty$

$$\int_0^{\infty} \frac{1 - \cos x}{x^2} dx = \frac{\pi}{2}.$$

The latter is a very important formula of integral calculus; in contrast to the preceding integral this integral is evidently *absolutely* convergent.

Let us make one more remark. If $\alpha > 0$, the change of the variable $z = x / \alpha$ gives:

$$\int_0^{\infty} \frac{\sin \alpha z}{z} dz = \int_0^{\infty} \frac{\sin x}{x} dx = \frac{\pi}{2};$$

if $\alpha < 0$ the replacement $z = -x / \alpha$ gives:

$$\int_0^{\infty} \frac{\sin \alpha z}{z} dz = - \int_0^{\infty} \frac{\sin x}{x} dx = - \frac{\pi}{2};$$

hence the integral

$$I(\alpha) = \int_0^{\infty} \frac{\sin \alpha z}{z} dz,$$

which converges for all values of α is a discontinuous function of α :

$$I(\alpha) = \begin{cases} +\frac{\pi}{2} & (\alpha > 0), \\ -\frac{\pi}{2} & (\alpha < 0), \\ 0 & (\alpha = 0). \end{cases}$$

Example 3. Let us now evaluate the integral

$$I = \int_0^{\infty} e^{-x^2} dx,$$

which plays an important part in many applications of integral calculus (for example in the theory of probability and mathematical statistics). The replacement of the variable $x = ut$ (where $u > 0$ is a constant) gives:

$$I = \int_0^{\infty} ue^{-u^2 t^2} dt \quad (u > 0).$$

Let us assume that for $u > 0$, $\varepsilon > 0$

$$f(u, t) = ue^{-u^2(1+t^2)},$$

$$\varphi(u) = \int_0^{\infty} f(u, t) dt = e^{-u} \int_0^{\infty} ue^{-u^2 t^2} dt = Ie^{-u^2}, \quad (1)$$

$$\psi_{\varepsilon}(t) = \int_{\varepsilon}^{\infty} f(u, t) du = -\frac{1}{2(1+t^2)} e^{-u^2(1+t^2)} \Big|_{\varepsilon}^{\infty} = \frac{e^{-\varepsilon^2(1+t^2)}}{2(1+t^2)}. \quad (2)$$

The function $\varphi(u)$ is continuous in the interval (ε, ∞) and the function $\psi_{\varepsilon}(t)$ in the interval $(0, \infty)$. The integral

$$\int_{\varepsilon}^{\infty} \varphi(u) du = I \int_{\varepsilon}^{\infty} e^{-u^2} du,$$

is evidently convergent. It therefore follows from theorem 4 § 110

$$\int_{\varepsilon}^{\infty} \varphi(u) du = \int_0^{\infty} \psi_{\varepsilon}(t) dt. \quad (3)$$

Let us take the limit of this equation for $\varepsilon \rightarrow 0$. As a result of (1) the left-hand side is equal to

$$I \int_{\varepsilon}^{\infty} e^{-u^2} du,$$

and it tends to I^2 as $\varepsilon \rightarrow 0$. In order to find the limit of the right-hand side we note that $t \geq 0$ for $\varepsilon \geq 0$ and, from (2),

$$\psi_{\varepsilon}(t) \leq \frac{1}{2(1+t^2)};$$

since the integral

$$\int_0^{\infty} \frac{dt}{2(1+t^2)}$$

is convergent, therefore, on the basis of the test in § 110, the integral on the right-hand side of the equation (3) is uniformly convergent in the region $\varepsilon \geq 0$ and is therefore a continuous function of ε in that region; in particular, its limit for $\varepsilon \rightarrow 0$ is equal to its value at $\varepsilon = 0$, i.e. it is equal to

$$\int_0^{\infty} \frac{dt}{2(1+t^2)} = \frac{1}{2} \arctan t \Big|_0^{\infty} = \frac{\pi}{4}.$$

We therefore obtain:

$$I^2 = \frac{\pi}{4}, \quad I = \frac{1}{2} \sqrt{\pi},$$

and our problem is solved.

§ 112. Euler's integrals

Euler's integrals are integrals

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$$

(integral of the first type) and

$$\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$$

(integral of the second type) *). The first integral is a function of two parameters p and q and the second a function of one parameter s . Both integrals define important non-elementary functions which are important in various applications. Their properties were therefore studied extensively; detailed tables were also composed for them. In this paragraph we shall study some of the simple properties of these functions.

1. Naturally both integrals only define the corresponding functions for those values of the parameters for which they are convergent. We must therefore begin by finding the regions of convergence for these integrals (*i.e.* the sets of values of the parameters for which they are convergent). In the integral $B(p, q)$ the integrand is continuous in the whole interval of integration when $p \geq 1$, $q \geq 1$ and there is no doubt as to the existence of the integral. But, when $p < 1$, the integrand evidently becomes unbounded in the neighbourhood of the point $x = 0$; when $q < 1$, the same takes place in the neighbourhood of the point $x = 1$. When $p < 1$, we have $(1 - x)^{q-1} \rightarrow 1$ as $x \rightarrow 0$ irrespective of q and therefore

$$x^{p-1} (1-x)^{q-1} \sim x^{p-1} \quad (x \rightarrow 0).$$

It follows from theorem 3' § 108 that the integral $B(p, q)$ will be divergent for $p \leq 0$ ($p - 1 \leq -1$); if, however, $p > 0$ ($p - 1 > -1$), then the behaviour of the integrand in the neighbourhood of the point $x = 0$ does not prevent the convergence of the integral. Similarly the behaviour of the integrand in the neighbourhood of the point $x = 1$ does not prevent convergence of the integral for $q > 0$, for this integral will only be divergent for $q \leq 0$. We can therefore conclude that *in order that the integral $B(p, q)$ should be convergent it is necessary and sufficient that $p > 0$, $q > 0$* and we shall assume this to be so in future. As we know from § 107, the infinite interval of integration in the integral $\Gamma(s)$ does not prevent convergence of the integral for any s . But if $s < 1$, the integrand is not bounded in the neighbourhood of the point $x = 0$; since for $x \rightarrow 0$

$$x^{s-1} e^{-x} \sim x^{s-1},$$

*) $B(p, q)$ and $\Gamma(s)$ are read as follows: "beta of p and q " and "gamma of s ".

therefore we can conclude that in order that the integral $\Gamma(s)$ should be convergent it is necessary and sufficient that $s > 0$.

2. Let us now show that the function $B(p, q)$ is continuous at every point of the region of convergence of the integral, i.e. for every value of $p > 0, q > 0$. This follows from the fact that for all arbitrary positive numbers p_0, q_0 the integral $B(p, q)$ will converge uniformly in the region $p \geq p_0, q \geq q_0$. In fact, if $p \geq p_0, q \geq q_0$, we have in the interval $(0, 1)$ irrespective of x

$$x^{p-1} (1-x)^{q-1} \leq x^{p_0-1} (1-x)^{q_0-1},$$

and therefore for every $\varepsilon > 0$

$$\int_0^\varepsilon x^{p-1} (1-x)^{q-1} dx \leq \int_0^\varepsilon x^{p_0-1} (1-x)^{q_0-1} dx \quad (p \geq p_0, q \geq q_0),$$

and we also have a similar inequality for the interval of integration $(1-\varepsilon, 1)$ which proves uniform convergence of the integral $B(p, q)$ in the region $p \geq p_0, q \geq q_0$ (cf. § 110 test for uniform convergence and final note).

Uniform convergence of the integral $\Gamma(s)$ is proved in exactly the same way in the region $s_0 \leq s \leq S_0$, where s_0 and S_0 ($s_0 < S_0$) are arbitrary positive numbers (for this integral the region of uniform convergence must be bounded from above unlike that for the integral $B(p, q)$, since it can readily be shown that if we have an infinite limit, convergence is non-uniform in the whole region $s \geq s_0$). The function $\Gamma(s)$ is therefore continuous for every $s > 0$.

3. When $s > 0$, integration by parts gives us:

$$\Gamma(s+1) = \int_0^\infty x^s e^{-x} dx = -x^s e^{-x} \Big|_0^\infty + s \int_0^\infty x^{s-1} e^{-x} dx = s \Gamma(s),$$

since $x^s e^{-x}$ vanishes at $x = 0$ and tends to zero as $x \rightarrow \infty$. Hence

$$\Gamma(s+1) = s \Gamma(s) \quad (s > 0). \quad (1)$$

Using the same relation again we obtain for every natural number n and for every $s > n-1$

$$\Gamma(s+1) = s(s-1) \dots (s-n+1) \Gamma(s-n+1). \quad (2)$$

In particular, when $s = n$

$$\Gamma(n+1) = n(n-1) \dots 2 \cdot 1 \cdot \Gamma(1)$$

But

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1,$$

and we obtain for integral values of $n \geq 1$

$$\Gamma(n+1) = n!$$

The importance of this remarkable formula is mainly due to the fact that it gives a simple analytical formula for the expression $n!$:

$$n! = \int_0^{\infty} x^n e^{-x} dx \quad (n = 1, 2, \dots), \quad (3)$$

which also hold when $n = 0$ if we assume as is usually done that $0! = 1$. But, in contrast to the expression for $n!$, the integral on the right-hand side of the equation (3) holds not only for integral values of n but for every $n > -1$; it is therefore useful to apply formula (3) for evaluating the expression for $n!$ when the values of $n > -1$ are fractional; as a result we obtain an analytical expression for $n!$ in the form of a continuous function of n which retains its usual value for integral values of $n \geq 0$.

If $s > 0$ is not an integer, then an integer $n \geq 0$ can be found such that $n < s < n+1$; according to formula (2) we have

$$\Gamma(s) = (s-1)(s-2) \dots (s-n)\Gamma(s-n);$$

since $0 < s-n < 1$, therefore, this formula replaces the study of the function $\Gamma(s)$ in the interval $(n, n+1)$ by the study of its behaviour in the interval $(0, 1)$; and since $n \geq 0$ is an arbitrary integer, therefore, by knowing the course of the function $\Gamma(s)$ in the interval $0 \leq s \leq 1$ we can evaluate it by means of simple elementary methods and determine its properties in the whole region $s > 0$. All this follows directly from the fundamental "functional equation" (1) which the function $\Gamma(s)$ satisfies.

4. Formulae for lowering the argument, analogous to formula (1), can be deduced for the function $B(p, q)$. Let $p > 0, q > 0$. Integration by parts give:

$$\begin{aligned} B(p+1, q+1) &= \int_0^1 x^p (1-x)^q dx = \\ &= \left[\frac{x^{p+1}}{p+1} (1-x)^q \right]_0^1 + \frac{q}{p+1} \int_0^1 x^{p+1} (1-x)^{q-1} dx; \end{aligned}$$

and since we have identically

$$x^{p+1} = x^p - x^p (1-x),$$

therefore

$$B(p+1, q+1) = \frac{q}{p+1} \{B(p+1, q) - B(p+1, q+1)\}$$

and consequently

$$B(p+1, q+1) = \frac{q}{p+q+1} B(p+1, q);$$

an analogous formula which lowers the first argument evidently also holds:

$$B(p+1, q+1) = \frac{p}{p+q+1} B(p, q+1).$$

By applying this last formula to the right-hand side of the previous formula we obtain the symmetrical formula:

$$B(p+1, q+1) = \frac{pq}{(p+q)(p+q+1)} B(p, q) \quad (p > 0, q > 0). \quad (3')$$

5. We shall now establish the remarkable connection between the functions $B(p, q)$ and $\Gamma(s)$. Replacing the variable of integration in the integral $B(p, q)$

$$x = \frac{1}{1+z},$$

we obtain

$$1-x = \frac{z}{1+z}, \quad dx = -\frac{dz}{(1+z)^2}, \quad z = \frac{1-x}{x},$$

and we readily find:

$$B(p, q) = \int_0^\infty \frac{z^{q-1}}{(1+z)^{p+q}} dz. \quad (4)$$

On the other hand, the transformation of the variable $x = z/\alpha$, where $\alpha > 0$ is constant, gives us for $s > 0$,

$$\int_0^{\infty} x^{s-1} e^{-\alpha x} dx = \frac{1}{\alpha^s} \int_0^{\infty} z^{s-1} e^{-z} dz = \frac{\Gamma(s)}{\alpha^s}; \quad (5)$$

therefore for $z > 0$, $p > 0$, $q > 0$

$$\int_0^{\infty} u^{p+q-1} e^{-u(1+z)} du = \frac{\Gamma(p+q)}{(1+z)^{p+q}}. \quad (6)$$

Formulae (4) and (6) give:

$$\begin{aligned} \Gamma(p+q) B(p, q) &= \int_0^{\infty} \Gamma(p+q) \frac{z^{q-1}}{(1+z)^{p+q}} dz = \\ &= \int_0^{\infty} \left\{ z^{q-1} \int_0^{\infty} u^{p+q-1} e^{-u(1+z)} du \right\} dz = \\ &= \int_0^{\infty} \left\{ \int_0^{\infty} z^{q-1} u^{p+q-1} e^{-u(1+z)} du \right\} dz. \end{aligned} \quad (7)$$

Let us at first assume that $p > 1$ and $q > 1$; then it can be readily shown that the order of integration on the right-hand side of this formula can be changed; we shall show in this connection that all assumptions made in theorem 4 § 110 also hold here. In fact, denoting the integrand by $f(z, u)$ we can readily see that this function is continuous and non-negative in the region of integration. Moreover, the function

$$\int_0^{\infty} f(z, u) du = z^{q-1} \int_0^{\infty} u^{p+q-1} e^{-u(1+z)} du = \Gamma(p+q) \frac{z^{q-1}}{(1+z)^{p+q}}$$

(cf. (6)) is continuous for $0 \leq z < +\infty$ and the function

$$\begin{aligned} \int_0^{\infty} f(z, u) dz &= e^{-u} u^{p+q-1} \int_0^{\infty} z^{q-1} e^{-uz} dz = e^{-u} u^{p+q-1} \frac{\Gamma(q)}{u^q} = \\ &= \Gamma(q) u^{p-1} e^{-u} \end{aligned} \quad (8)$$

(we have used formula (5) for integration) is continuous for $0 \leq u < +\infty$. In accordance with theorem 4 § 110 it therefore follows that existence of the integral on the right-hand side of (7) implies convergence of the integral

$$\int_0^{\infty} \left\{ \int_0^{\infty} f(z, u) dz \right\} du$$

and that these two integrals are equal to one another so that, according to formula (8)

$$\begin{aligned} \Gamma(p+q) B(p, q) &= \int_0^{\infty} \left\{ \int_0^{\infty} f(z, u) dz \right\} du = \\ &= \int_0^{\infty} \Gamma(q) u^{q-1} e^{-u} du = \Gamma(p) \Gamma(q). \end{aligned}$$

Therefore for all values of $p > 1, q > 1$ we have:

$$B(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)}. \quad (9)$$

Let us now assume that p and q are two arbitrary positive numbers. It therefore follows from the above proof that we have (since $p+1 > 1, q+1 > 1$);

$$B(p+1, q+1) = \frac{\Gamma(p+1) \Gamma(q+1)}{\Gamma(p+q+2)}.$$

Expressing all values of the functions B and Γ by means of the formulae (1) and (3') we obtain formula (9) after making obvious reductions; hence formula (9) is proved for all values of $p > 0, q > 0$.

This most important relation thus enables us to replace the study of the function $B(p, q)$ by the study of the function $\Gamma(s)$ and in some cases, conversely, to elucidate the properties of the function $\Gamma(s)$ from the corresponding properties of the function $B(p, q)$.

In particular, when p and q are natural numbers, it follows from formula (9) that

$$B(p, q) = \frac{(p-1)!(q-1)!}{(p+q-1)!}.$$

Example 1. In many applications it is useful to evaluate the formula

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} x^{-\frac{1}{2}} e^{-x} dx.$$

The replacement of the variable of integration $x = u^2$ gives:

$$\Gamma\left(\frac{1}{2}\right) = 2 \int_0^{\infty} e^{-u^2} du = \sqrt{\pi} \quad (10)$$

(cf. example 3 § 111).

Example 2. The theory of Euler's integrals enables us to evaluate integrals of the following type which occur frequently in applications:

$$A_{nm} = \int_0^{\frac{\pi}{2}} \sin^n x \cos^m x dx,$$

where n and m are non-negative integers. The replacement of the variable $\sin^2 x = u$ gives:

$$\begin{aligned} A_{nm} &= \frac{1}{2} \int_0^1 u^{\frac{n-1}{2}} (1-u)^{\frac{m-1}{2}} du = \frac{1}{2} B\left(\frac{n+1}{2}, \frac{m+1}{2}\right) = \\ &= \frac{1}{2} \frac{\Gamma\left(\frac{n+1}{2}\right) \Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{n+m}{2} + 1\right)}. \end{aligned}$$

In this expression all arguments of the function Γ are either integers or halves of integers. Hence by using formula (10) all the three values of this function can be evaluated in every case.

§ 113. Stirling's formula

In § 112 we have obtained the formula

$$n! = \int_0^{\infty} x^n e^{-x} dx \quad (1)$$

which, when $n \geq 0$ is an integer, gives a simple analytical expression for $n!$ and defines the function $n!$ for other values of $n \geq -1$. The factorials of large numbers frequently play an important part both in theoretical considerations and in practical calculations; and since, by definition, the factorial of a large number has a complicated structure which is inconvenient for evaluations whose accuracy cannot be determined directly, therefore for large values of n it is convenient to have a simple readily accessible analytical expression for $n!$. Formula (1) does not by itself provide this expression, since it is not sufficiently clear for direct evaluations. However, by taking formula (1) as a basis we can deduce an approximate formula for $n!$ which satisfies all these requirements. This paragraph is devoted to this deduction which is very instructive; the method for the evaluation of the integral (1) which lies at its basis is frequently also used in other analytical problems.

For greater clarity we shall divide the deduction into several stages.

1. We shall at first convert the integral (1) into a more convenient form by replacing the variable of integration

$$x = n(1 + z), \quad dx = ndz, \quad z = \frac{x - n}{n},$$

which gives :

$$n! = n^{n+1} e^{-n} \int_{-1}^{\infty} \{ (1 + z) e^{-z} \}^n dz;$$

we shall assume in all what follows that :

$$(1 + z) e^{-z} = \varphi(z),$$

so that

$$n! = n^{n+1} e^{-n} \int_{-1}^{\infty} \{ \varphi(z) \}^n dz. \quad (2)$$

2. In order to study the behaviour of the integrand we must now study the elementary function $\varphi(z)$ in detail. Since

$$\varphi'(z) = -ze^{-z},$$

therefore the function $\varphi(z)$ increases for $z < 0$ and decreases for $z > 0$; it has its maximum equal to unity at the point $z = 0$; $\varphi(-1) = 0$ and when $z > 0$, the function $\varphi(z)$ is positive and monotonically tends to zero as $z \rightarrow \infty$; the graph of this function is schematically represented in fig. 69.

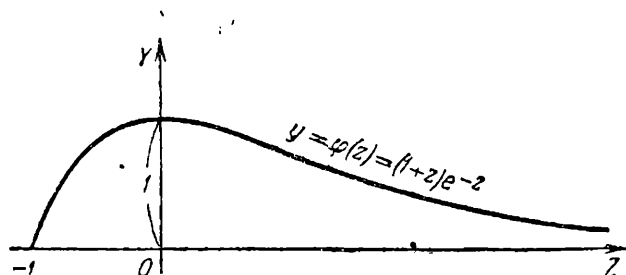


Fig. 69

We have further

$$\ln \varphi(z) = \ln(1+z) - z.$$

According to Taylor's formula

$$\ln(1+z) = z - \frac{z^2}{2} + \alpha z^3,$$

where α depends on z and it tends to a definite limit for $z \rightarrow 0$ and is therefore bounded. Hence

$$\ln \varphi(z) = -\frac{z^2}{2} + \alpha z^3,$$

and consequently

$$n \ln \varphi(z) = -\frac{nz^2}{2} + \alpha n z^3,$$

and therefore

$$\{\varphi(z)\}^n = e^{-\frac{nz^2}{2}} e^{\alpha n z^3}. \quad (3)$$

Let us note that the infinitely small quantity $e^t - 1$ is equivalent to t for $t \rightarrow 0$ i. e. the ratio $(e^t - 1)/t$ tends to unity as $t \rightarrow 0$ so that

$$e^t = 1 + \beta t,$$

where β remains bounded for $t \rightarrow 0$. Therefore if n and z change such that $nz^3 \rightarrow 0$, then

$$e^{\alpha n z^3} = 1 + \beta \alpha n z^3 = 1 + \gamma n z^3,$$

where γ is bounded for $nz^3 \rightarrow 0$; hence the equation (3) gives us for $nz^3 \rightarrow 0$:

$$\{ \varphi(z) \}^n = e^{-\frac{nz^2}{2}} \{ 1 + \gamma nz^3 \},$$

where γ is bounded. This equation gives us the required evaluation for the integrand in the integral (2).

3. Let us now make a plan for further evaluation of the integral

$$I(n) = \int_{-1}^{\infty} \{ \varphi(z) \}^n dz. \quad (4)$$

Since the function $\varphi(z)$ is equal to unity at $z = 0$ and it is confined between 0 and 1 at all other points of the interval of integration (cf. Fig. 69), therefore, for large values of n the integrand is negligibly small everywhere except in a short interval which surrounds the point $z = 0$; the region in which the integrand has somewhat greater values will be the smaller, the larger n is. This naturally leads us to believe that we should exclude from the interval of integration a short interval $(-\lambda, \lambda)$ surrounding the point $z = 0$; this must be done enough accurately taking into consideration the smallness of the values of z involved in the integration; we must evaluate the integral along that interval and subsequently show, by using rougher approximations, that the integrals along other interval of the general interval of integration are negligibly small in comparison to the calculated value along the interval $(-\lambda, \lambda)$. It is obvious that the number λ must, in this case, depend on n and tend to zero for $n \rightarrow \infty$.

This is the method which we shall use. Let θ denote an arbitrary constant confined between $1/3$ and $1/2$ (for example $\theta = 2/5$ or $\theta = 5/12$). We find the calculation that a convenient value for λ is

$$\lambda = n^{-\theta}.$$

We shall use this value in further arguments. We thus divide the interval of integration $(-1, +\infty)$ into three parts: $(-1, -\lambda)$, $(-\lambda, \lambda)$ and $(\lambda, +\infty)$; the integral (4) is correspondingly broken up into three terms each of which will be evaluated separately.

4. Let us begin with the most important integral

$$I_1(n) = \int_{-\lambda}^{\lambda} \{ \varphi(z) \}^n dz.$$

We have along the whole interval of integration :

$$|nz^3| \leq n\lambda^3 = n^{1-3\theta} \rightarrow 0 \quad (n \rightarrow \infty),$$

since $\theta > 1/3$, and hence $1 - 3\theta < 0$. We can therefore conclude from 2 that in the whole interval $(-\lambda, \lambda)$

$$\{\varphi(z)\}^n = e^{-\frac{nz^2}{2}} \{1 + \gamma nz^3\},$$

where γ remains uniformly bounded in the interval $(-\lambda, \lambda)$ for $n \rightarrow \infty$:

$$|\gamma| < c \quad (n \rightarrow \infty, |z| \leq \lambda),$$

where c is a positive constant. Therefore

$$\left| I_1(n) - \int_{-\lambda}^{\lambda} e^{-\frac{nz^2}{2}} dz \right| \leq cn\lambda^3 \int_{-\lambda}^{\lambda} e^{-\frac{nz^2}{2}} dz < cn^{1-3\theta} \int_{-\infty}^{+\infty} e^{-\frac{nz^2}{2}} dz.$$

But

$$\int_{-\infty}^{+\infty} e^{-\frac{nz^2}{2}} dz = \sqrt{\frac{2}{n}} \int_{-\infty}^{+\infty} e^{-u^2} du = 2 \sqrt{\frac{2}{n}} \int_0^{\infty} e^{-u^2} du = \frac{\sqrt{2\pi}}{\sqrt{n}}$$

(cf. § 111, example 3); therefore when $n \rightarrow \infty$,

$$\left| I_1(n) - \int_{-\lambda}^{\lambda} e^{-\frac{nz^2}{2}} dz \right| < cn^{1-3\theta} \frac{\sqrt{2\pi}}{\sqrt{n}} = o\left(\frac{1}{\sqrt{n}}\right). \quad (3)$$

The integral

$$\int_{-\lambda}^{\lambda} e^{-\frac{nz^2}{2}} dz$$

differs from the integral evaluated above

$$\int_{-\infty}^{+\infty} e^{-\frac{nz^2}{2}} dz = \frac{\sqrt{2\pi}}{\sqrt{n}}$$

only by the double value of the integral

$$\int_{\lambda}^{\infty} e^{-\frac{nz^2}{2}} dz = \frac{\sqrt{2}}{\sqrt{n}} \int_{\lambda \sqrt{\frac{n}{2}}}^{\infty} e^{-u^2} du;$$

for the lower limit of the last integral is equal to $n^{\frac{1}{2}-\theta} / \sqrt{2}$, since $\lambda = n^{-\theta}$ and, since $\theta < 1/2$, it increases indefinitely together with n ; therefore the last integral is indefinitely small for $n \rightarrow \infty$; the whole right-hand side of the last equation is a quantity of the form $o(1/\sqrt{n})$. Hence the relation (5) gives :

$$I_1(n) - \frac{\sqrt{2\pi}}{\sqrt{n}} < 2 \int_{\lambda}^{\infty} e^{-\frac{nz^2}{2}} dz + o\left(\frac{1}{\sqrt{n}}\right) = o\left(\frac{1}{\sqrt{n}}\right),$$

or

$$I_1(n) = \frac{\sqrt{2\pi}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right). \quad (6)$$

We have thus reached our goal as far as the integral $I_1(n)$ is concerned : we represent it as a sum of a very simple "principal term" $\sqrt{2\pi}/\sqrt{n}$ and another term about which we only know that it is infinitely small in comparison to the principal term for $n \rightarrow \infty$.

5. Let us now evaluate the integral

$$I_2(n) = \int_{-1}^{-\lambda} \{ \varphi(z) \}^n dz.$$

Since the function $\varphi(z)$ increases in the interval of integration, therefore :

$$I_2(n) < \{ \varphi(-\lambda) \}^n (1 - \lambda) < \{ \varphi(-\lambda) \}^n; \quad (7)$$

we

$$\varphi(-\lambda) = (1 - \lambda) e^{\lambda},$$

$$\ln \varphi(-\lambda) = \lambda + \ln(1 - \lambda) = \lambda - \left[\lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3} + \dots \right] < -\frac{\lambda^2}{2},$$

$$n \ln \varphi(-\lambda) < -\frac{n\lambda^2}{2} = -\frac{n^{1-2\theta}}{2},$$

$$\{ \varphi(-\lambda) \}^n < e^{-\frac{1}{2} n^{1-2\theta}} = e^{-\frac{1}{2} n^{\epsilon}},$$

where $\tau = 1 - 2\theta > 0$. But we know (cf. example 7 at the end of § 37) that for arbitrarily small $\tau > 0$ the right-hand side of the last equation is infinitely small in comparison to any power of n for $n \rightarrow \infty$,

for example in comparison with $\frac{1}{\sqrt{n}} = n^{-\frac{1}{2}}$. Therefore equation (7) gives :

$$I_2(n) = o\left(\frac{1}{\sqrt{n}}\right). \quad (8)$$

6. Let us finally consider the integral

$$I_3(n) = \int_{\lambda}^{\infty} \{ \varphi(z) \}^n dz.$$

Here it is again convenient to divide the interval of integration into two parts :

$$I_3(n) = I'_3(n) + I''_3(n) = \int_{\lambda}^4 \{ \varphi(z) \}^n dz + \int_4^{\infty} \{ \varphi(z) \}^n dz.$$

The function $\varphi(z)$ decreases in the interval $(\lambda, 4)$ and therefore does not exceed anywhere the value $\varphi(\lambda)$ so that we have :

$$I'_3(n) = \int_{\lambda}^4 \{ \varphi(z) \}^n dz \leq \{ \varphi(\lambda) \}^n (4 - \lambda) < 4 \{ \varphi(\lambda) \}^n. \quad (9)$$

But by means of Taylor's formula we readily obtain for n

$$\ln \varphi(\lambda) = -\frac{\lambda^2}{2} + o(\lambda^2),$$

and therefore for a sufficiently large value of n

$$\ln \varphi(\lambda) < -\frac{\lambda^2}{3},$$

hence

$$\{ \varphi(\lambda) \}^n < e^{-\frac{n\lambda^2}{3}} = e^{-\frac{n^{1-2\theta}}{3}} = e^{-\frac{n^{\tau}}{3}}$$

where $\tau = 1 - 2\theta > 0$. As in 5 we can therefore conclude from (9) that for $n \rightarrow \infty$

$$I'_3(n) = o\left(\frac{1}{\sqrt{n}}\right). \quad (10)$$

In order to evaluate the integral $I''_3(n)$ we note that for $x \geq 4$ as can be readily calculated*), $1+x < e^{+x/2}$, and therefore $\varphi(x) < e^{-x/2}$. Hence

$$I''_3(n) = \int_4^\infty \{\varphi(z)\}^n dz < \int_4^\infty e^{-\frac{nx}{2}} dx < \int_0^\infty e^{-\frac{nx}{2}} dx = \frac{2}{n} = o\left(\frac{1}{\sqrt{n}}\right) \quad (11)$$

for $n \rightarrow \infty$. It follows from (10) and (11) that

$$I_3(n) = I'_3(n) + I''_3(n) = o\left(\frac{1}{\sqrt{n}}\right). \quad (12)$$

7. The relations (6), (8) and (12) give :

$$\begin{aligned} I(n) &= \int_{-1}^\infty \{\varphi(z)\}^n dz = I_1(n) + I_2(n) + I_3(n) = \frac{\sqrt{2\pi}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) = \\ &= \frac{\sqrt{2\pi}}{\sqrt{n}} [1 + o(1)]. \end{aligned}$$

We therefore have as a result of (2) :

$$n! = n^{n+1} e^{-n} \frac{\sqrt{2\pi}}{\sqrt{n}} [1 + o(1)] = n^n e^{-n} \sqrt{2\pi n} [1 + o(1)].$$

We have thus achieved our purpose. The composite function $n!$ can be approximately represented by the simple and convenient analytical formula

$$\Gamma(n+1) = n! = n^n e^{-n} \sqrt{2\pi n} [1 + o(1)].$$

The latter is one of the most important formulae of mathematical analysis and is usually known as *Stirling's formula* ; it has numerous applications. With regard to the last term of this formula we have only established that it is infinitely small in comparison with the

*) It is sufficient to show that the function $e^{x/2} - 1 - x$ is positive at $x = 4$ and increases for $x \geq 4$.

principal term of this formula for $n \rightarrow \infty$; in other words, we only aim at isolating the principal term of *Stirling's* formula and we are not concerned with the evaluation of the last term. In practice the latter evaluations are often also needed; they can be obtained without undue difficulties by using the method used for obtaining the expression for the principal term; however, for this purpose all calculations would have to be performed with much greater accuracy and we cannot go into detail within the scope of this book.

Stirling's formula is frequently written in the equivalent logarithmic form

$$\ln \Gamma(n+1) = \ln(n!) = n \ln n - n + \frac{1}{2} \ln n + \frac{1}{2} \ln 2\pi + o(1).$$

CHAPTER XXVII

DOUBLE AND TRIPLE INTEGRALS

§ 114. Measurable plane figures

1. Introduction. In the same way as the methods and concepts of differential calculus, initially considered for functions of one independent variable, were subsequently developed for functions of any number of variables (chapter 22), so the fundamental concepts of integral calculus can also be successfully extended to functions of several variables. This holds for the basic concept of an integral as limit of a sum of a definite form; we have seen that the origin of this concept was mainly due to practical requirements. In this chapter we shall briefly consider fundamental problems connected with integration of functions of two or three variables. All that we shall say in this connection can be easily extended to functions of any number of variables.

The region of integration of a function of one variable usually consists of an interval or, in more complicated cases, a group of intervals. The region of integration of a function of two variables is a plane figure. If we only consider the simplest figures, *i.e.* figures bounded by simple closed curves, we must still deal with a large variety of forms of such "regions of integration". This variety which has no parallel with functions of one variable requires a detailed study of some properties of plane figures before we can proceed with the study of double integrals. This paragraph is devoted to this subject.

2. Plane figures. In its most general form we shall call a plane figure an arbitrary set F of points in a plane. An arbitrary point A in the plane is said to be an *interior point* of the figure F if it is the centre of a circle with its all points belonging to the figure F (fig. 70). The point B with centre of a circle not having any point which

belongs to the figure F is said to be an *exterior point* of that figure. Finally a point in the plane which is neither an interior nor an exterior point of the figure F is called a *boundary point*. Thus the point C is evidently characterised by the fact that any circle with centre at C contains points which belong to the figure F as well as other points which do not belong to that figure.

It follows from the definition that an interior point of the figure F always belongs to that figure whereas an exterior point can never belong to it. A boundary point may or may not belong to the figure F . Thus if F is the set of interior points of a circle, then no point on the boundary of the figure F belongs to this circle; if however F represents the set of interior points of a circle as well as points on its circumference, then such a figure will contain all its boundary points.

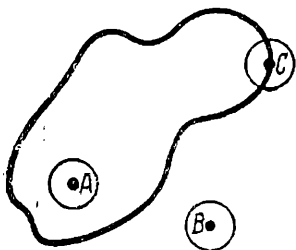


Fig. 70

The set of all boundary points of a given figure (*i.e.* points which belong and which do not belong to this figure) is called the *contour* (or *boundary*) of this figure.

A figure which does not contain a single boundary point (*i.e.* all points are interior points) is said to be an *open region*; a figure which contains all its boundary points is said to be a *closed region*. If M , N , P respectively denote the sets of all interior, exterior and boundary points of a figure F , then, as the reader can readily see, M and N are always open regions whereas P , $M+P$ and $N+P$ are always closed regions.

Lemma 1. *The rectilinear section AB which connects the interior point A with the exterior point B of the figure F contains at least one boundary point of that figure.*

Proof. Let us agree to call every rectilinear section *normal* if one of its ends is an interior and the other an exterior point of the figure F . According to the conditions of this lemma the section AB is normal; if its middle does not lie on the boundary of the figure F , then, evidently, one of the two halves of this section will be normal. To this half we can again apply the above argument, and so on. Sooner or later we shall obtain a section whose centre lies on the boundary of the figure F (and this will prove our lemma) or we may obtain a contracting sequence of normal sections. In that event the common point D of all these sections (lemma 1 § 18) evidently

possesses the property that any circle with centre at D contains an infinite number of normal sections, *i.e.* it contains points which belong to the figure F as well as points which do not belong to it. But this implies that D lies on the boundary of the figure F . Our lemma is thus proved.

3. Measure of plane figures. In chapter 12 we have determined the area of curvilinear trapeziums and figures consisting of curvilinear trapeziums. We must now consider the evaluation of areas on a wider basis. In order to distinguish our new definition of areas from the former definition we shall now speak not of areas but of *measure* of plane figures. In doing this we must try to generalise our new definition and see that this measure should, in fact, always exist and coincide with the area defined earlier for figures whose areas can be calculated in accordance with our former definition.

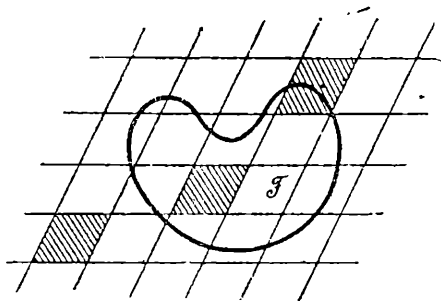


Fig. 71

Let us assume that we are given an arbitrary bounded plane figure F . Let us draw in the given plane two families of mutually parallel lines as shown in fig. 71. The lines of each family are such that the distance between any two adjacent lines is always constant. The drawn net of straight lines evidently divides our plane into equal parallelograms which we shall call the *cells* of our net; we shall always regard all points on the contour of such a parallelogram as belonging to the corresponding cell, so that each cell is a closed region. The longest diagonal of the parallelogram is evidently the diameter of this cell. This quantity, which is the same for all cells, is called *parameter* of the given net S and denoted by $\rho(S)$ or simply by ρ .

The cells of the net S can, in general, be divided into two categories with reference to the figure F : *inner* cells, all points of which are interior points of the figure F and *outer* cells, all points of which are exterior points of the figure F ; finally all remaining cells are called *boundary* cells (with reference to the figure F ; in fig. 71 one cell of each type is shaded).

Lemma 2. *Each boundary cell contains at least one point on the boundary of the figure F .*

Proof. In fact, otherwise each point of the given cell would either be an interior or an exterior point (with reference to F) and not all points can then be interior or exterior points, for if it were so, then the cell would either be an inner or an outer cell. Therefore the given cell must contain the interior point A and the exterior point B of the figure F . In accordance with lemma 1 the section AB which completely belongs to the given section contains a point on the boundary which proves lemma 2.

• Let $I_s(F)$ denote the sum of areas of all inner cells of the net S (with reference to the figure F). Evidently the sum of these areas is different for different nets S ; however, the set of the values of the quantity $I_s(F)$ has evidently an upper limit for all possible types of nets, since the figure F which by definition is bounded lies entirely within a circle C and the quantity $I_s(F)$ can never exceed the area of this circle irrespective of S . Therefore there exists an upper bound $I(F)$ for all the quantities $I_s(F)$.

Theorem 1, $I_s(F) \rightarrow I(F)$ for $\rho(S) \rightarrow 0$.

As usual, the statement of theorem 1 implies as follows: There exists a $\delta > 0$ for an arbitrary $\varepsilon > 0$ such that $I_s(F) > I(F) - \varepsilon$ for every net S with a parameter $\rho < \delta$.

Proof. Since $I(F)$ is the upper bound of the numbers $I_s(F)$, therefore for every $\varepsilon > 0$ a net S_0 can be found such that

$$I_{s_0}(F) > I(F) - \varepsilon.$$

We shall regard this net as fixed in all subsequent arguments of this proof. Let us consider an arbitrary inner cell Δ of the net S_0 . This cell is a closed region; on the other hand, the contour D of the figure F is, as we know, also a closed region. The regions Δ and D have no points in common, since all points of the cell Δ are interior points of the figure F . It follows from theorem 4 § 87 that the distance between the regions Δ and D is positive. This argument holds for every inner cell Δ of the net S_0 . If δ denotes the shortest distance obtained in this way, then any point of any inner cell of the net S_0 is evidently at a distance not less than δ from the contour of the figure F .

Let S now denote an arbitrary net in which $\rho(S) < \delta$ and let A be an arbitrary point of an inner cell of the net S_0 . A is an interior point of the figure F and therefore it must belong to an inner or a boundary cell of the net S ; however, it cannot belong to a boundary

cell, since in that case, in accordance with lemma 2, it would be at a distance less than δ from the contour of the figure F . Hence every point A of every inner cell of the net S_0 must belong to an inner cell of the net S . Hence

$$I_s(F) \geq I_{s_0}(F) > I(F) - \varepsilon,$$

provided $\rho < \delta$ for the net S . This proves theorem 1.

The set of inner cells of any net of straight lines evidently belongs entirely to the figure F . If we add the boundary cells to inner cells of the same net, the sum of whose areas being denoted by $K_s(F)$, we evidently obtain a figure which, conversely, contains all points of the figure F . It is thus evident that if we can ascribe a definite area to the figure F , then for every net S this area must be confined between $I_s(F)$ and $I_s(F) + K_s(F)$. But we have just proved that $I_s(F)$ always tends to a definite limit as $\rho \rightarrow 0$. Hence the area $I_s(F) + K_s(F)$ which surrounds the figure F tends to the same limit; we are therefore justified in assuming that this limit is equal to the area or, as we now prefer to say, to the *measure* of the figure F ; the figure F itself is in this case said to be *measurable*. Since $I_s(F) \rightarrow I(F)$ always (theorem 1), therefore in order that the figure F should be measurable it is necessary and sufficient that

$$K_s(F) \rightarrow 0 \quad (\rho \rightarrow 0),$$

i.e. the sum of the areas of the boundary cells should tend to zero together with the parameter of the net.

In this case the above defined quantity $I(F)$ is the measure of the figure F .

4. Properties of measurable figures. The sum $F_1 + F_2$ of two measurable figures F_1 and F_2 is the set of points which belong to at least one of these figures; the sum of any number of figures is determined similarly. The *intersection* or *common part* $F_1 F_2$ of the figures F_1 and F_2 is the set of points which belong both to F_1 and F_2 ; this definition also applies to the intersection of any number of figures.

Theorem 2. *If the figures F_1 and F_2 are measurable, then the figure $F_1 + F_2$ is also measurable; if at the same time the figures F_1 and F_2 have no common interior points, then*

$$I(F_1 + F_2) = I(F_1) + I(F_2).$$

Proof. 1) Every interior point of one of the figures F_1 and F_2 is also an interior point of the figure $F_1 + F_2$. Therefore point A

on the boundary of the figure $F_1 + F_2$ cannot be an interior point of either F_1 or F_2 . However, it cannot be an exterior point of either figures, for then it would evidently be an exterior point of the figure $F_1 + F_2$. Hence the point A must be a boundary point of at least one of the figures F_1 and F_2 . But in this case every cell of any net S which is a boundary cell of the figure $F_1 + F_2$ must also be a cell on the boundary of either F_1 or F_2 ; we therefore have for every net S :

$$K_s(F_1 + F_2) \leq K_s(F_1) + K_s(F_2).$$

But since the figures F_1 and F_2 are measurable, therefore $K_s(F_1)$ and $K_s(F_2)$ tend to zero as $\rho \rightarrow 0$; hence the last inequality shows that we also have

$$K_s(F_1 + F_2) \rightarrow 0 \quad (\rho \rightarrow 0),$$

and this, in its turn, implies that the figure $F_1 + F_2$ is measurable.

2) Let us now assume that the figures F_1 and F_2 have no interior points in common. Let us consider an arbitrary inner cell Δ of an arbitrary net S with reference to the figure $F_1 + F_2$. It is evident that the cell Δ cannot be an outer cell of both the figures F_1 and F_2 , for in that case it would also be an outer cell of the figure $F_1 + F_2$. Therefore the cell Δ must either be an inner cell or a boundary cell of at least one of the figures F_1 and F_2 and, consequently

$$I_s(F_1 + F_2) \leq I_s(F_1) + K_s(F_1) + I_s(F_2) + K_s(F_2). \quad (1)$$

Since F_1 and F_2 have no interior points in common, the set of cells which are inner cells of either F_1 or F_2 have an area equal to $I_s(F_1) + I_s(F_2)$; and since each of these cells is also an inner cell of the figure $F_1 + F_2$, therefore

$$I_s(F_1) + I_s(F_2) \leq I_s(F_1 + F_2). \quad (2)$$

It follows from (1) and (2)

$$I_s(F_1) + I_s(F_2) \leq I_s(F_1 + F_2) \leq I_s(F_1) + K_s(F_1) + I_s(F_2) + K_s(F_2).$$

By making the net S successively small we note that

$$I_s(F_1) \rightarrow I(F_1), \quad I_s(F_2) \rightarrow I(F_2), \quad I_s(F_1 + F_2) \rightarrow I(F_1 + F_2), \\ K_s(F_1) \rightarrow 0, \quad K_s(F_2) \rightarrow 0,$$

and we obtain in the limit:

$$I(F_1 + F_2) = I(F_1) + I(F_2),$$

which was to be proved.

It is evident that theorem 2 can be extended to include any number of terms by simple induction.

Theorem 3. *The intersection F_1F_2 of two measurable figures is also measurable.*

Proof. Let A be an arbitrary point on the boundary of the figure F_1F_2 ; it can be readily seen that in this case the point A must be a boundary point of at least one of the figures F_1 and F_2 ; in fact, the point A cannot be an exterior point of either F_1 or F_2 , for it would be an exterior point of the figure F_1F_2 as well; on the other hand, it cannot be an interior point of either figure, for in that case it would also be an interior point of the figure F_1F_2 . Therefore the point A on the boundary of the figure F_1F_2 must be a boundary point of either F_1 or F_2 ; hence for any net S a boundary cell with respect to F_1F_2 must be a boundary cell either with respect to F_1 or with respect to F_2 ; this gives:

$$K_s(F_1F_2) \leq K_s(F_1) + K_s(F_2).$$

Since F_1 and F_2 are measurable, the right-hand side tends to zero as $\varphi \rightarrow 0$; the same also holds for the left-hand side and this implies that the figure F_1F_2 is measurable.

Theorem 3 evidently holds for the intersection of any number of measurable figures.

Theorem 4. *Let $A(\delta)$ denote a set of points in a plane, which are at a distance less than δ from the contour of the plane figure F . Therefore in order that the figure F should be measurable it is necessary and sufficient that the set $A(\delta)$ should be contained in a finite group of parallelograms, the sum of whose areas is less than ε for arbitrarily small $\varepsilon > 0$ and sufficiently small δ .*

Proof. 1) *Necessity.* Let the figure F be measurable and let S be a net of straight lines for which φ is so small that

$$K_s(F) < \varepsilon;$$

let δ be the lower bound of the distances of all inner and outer cells of the net S from the contour of the figure F ($\delta > 0$ as a result of theorem 4 § 87). The points of the set $A(\delta)$ cannot therefore belong to inner or outer cells of the net S so that the set $A(\delta)$ must be contained entirely by boundary cells of this net, the sum $K_s(F)$ of whose areas is less than ε .

2) *Sufficiency.* If we have $\rho < \delta$ for the net S , then all boundary cells of this net belong entirely to the set $A(\delta)$; therefore if the condition of theorem 4 is satisfied, then for arbitrarily small $\varepsilon > 0$ we have for a sufficiently small ρ

$$K_s(F) < \varepsilon,$$

which implies that the figure F is measurable.

Theorem 4 has many important corollaries. Let us assume that the measurable figure F is divided into parts of arbitrary shape which we shall, as before, call cells, and for which we shall only require that they should be measurable figures with no interior points in common in pairs. Let $\rho(T)$ denote the greatest diameter of the cell of the division T of the figure F . The definition of inner and boundary cells remains as before (*i.e.*, we say that a cell is an inner cell if all its points are interior points of the figure F , otherwise it is a boundary cell). Let us denote by $I_T(F)$ the sum of measures of the inner cells of the division T (with reference to the figure F) and by $I(F)$, as before, the measure of the figure F .

Theorem 5. *If the figure F is measurable and can be divided into measurable cells, then for $\rho \rightarrow 0$*

$$I_T(F) \rightarrow I(F).$$

In fact, this implies as follows: a $\delta > 0$ can be found for every $\varepsilon > 0$ such that, we have for $\rho(T) < \delta$:

$$|I_T(F) - I(F)| < \varepsilon.$$

Proof. If $\rho(T) < \delta$, every point of any boundary cell of the division T will belong to the set $A(\delta)$ of theorem 4 and therefore, provided δ is sufficiently small, the set of all boundary cells will be contained by a finite group of parallelograms, the sum of whose areas is less than ε . Hence the sum of the areas of the boundary cells is less than ε *). But the sum of measures of all cells (inner and boundary cells) is equal to $I(F)$ (theorem 2); therefore the sum $I_T(F)$ of measures of inner cells differs from $I(F)$ by less than ε , which proves theorem 5.

*) Strictly speaking, in order to prove this statement we must also show that : 1) if the measurable figure F_1 contains a part of the measurable figure F_2 , then $I(F_1) \leq I(F_2)$ and 2) the measure of a parallelogram is equal to its area. This first statement follows directly from the definition of measure and the reader will have no difficulty in proving it. The second statement, which is almost self-evident, will be proved somewhat later (theorem 6).

5. Examples of measurable figures. We shall now show that there exist wide classes of more or less simple measurable figures and that for figures, whose areas are calculated by means of other special methods, their area coincides with the measure.

Lemma 3. *Let the function $y = f(x)$ be continuous in the interval (a, b) and let for every $\delta > 0$ $A(\delta)$ be a set of points in the plane, which are at a distance less than δ from the curve $y = f(x)$ ($a \leq x \leq b$). In that case for arbitrarily small $\varepsilon > 0$ and sufficiently small δ the set $A(\delta)$ can be covered by a finite group of rectangles the sum of whose areas is less than ε .*

Proof. Let us divide the interval (a, b) into n equal parts $(b - a) / n = \delta$ in length and let n be so large that the vibration of the function $f(x)$ in any subinterval $\leq \delta$ in length should not exceed ε . Let the points of division be $a = x_0 < x_1 < \dots < x_n = b$ so that $x_k - x_{k-1} = \delta$ ($1 \leq k \leq n$). Let us denote by M_k and m_k the greatest and smallest value of the function $f(x)$ in the subinterval (x_{k-1}, x_k) of length equal to δ , so that $M_k - m_k \leq \varepsilon$ (fig. 72).

Let MN denote the interval $x_{k-1} \leq x \leq x_k$ of the curve $y = f(x)$. It is evident that every point in the plane which is at a distance less than δ from the arc MN must lie 1) within the strip $x_{k-1} - \delta < x < x_k + \delta$ and 2) within the strip $m_k - \delta < y < M_k + \delta$ and hence within the rectangle $(x_{k-1} - \delta < x < x_k + \delta, m_k - \delta < y < M_k + \delta)$ whose area is

$$3\delta (M_k - m_k + 2\delta) \leq 3\delta (\varepsilon + 2\delta);$$

the sum of the areas of these rectangles constructed for all k ($1 \leq k \leq n$) does not exceed $3\delta (\varepsilon + 2\delta) n = 3(b - a)(\varepsilon + 2\delta)$ and is therefore as small as we please, provided ε is sufficiently small and n sufficiently large. And since the set of these rectangles covers the set $A(\delta)$, therefore lemma 3 is proved.

Let us now assume that the contour of the figure F is closed and can be divided into a finite number of intervals, in each of which the curve can be represented by either the equation $y = f(x)$ or the equation $x = \varphi(y)$. Applying lemma 3 to each interval we see that the set $A(\delta)$ constructed for the whole contour of the figure F can be covered by a finite group of rectangles, the sum of whose areas can be as small as we please provided δ is sufficiently small. Theorem 4 thus gives us the following theorem:

Theorem 6. *If the contour of the figure F can be divided into a finite number of sections each of which can be expressed by either of the two equa-*

tions : $y=f(x)$, $x=\varphi(y)$ (where f and φ are continuous functions), then the figure F is measurable.

By the way all trapeziums whose areas have been evaluated in chapter 12 will be measurable; for a trapezium its measure coincides with its area as determined by the methods described in chapter 12. Measurability of every rectangle follows directly from theorem 6. Hence every rectangle is measurable and in order to find its mea-

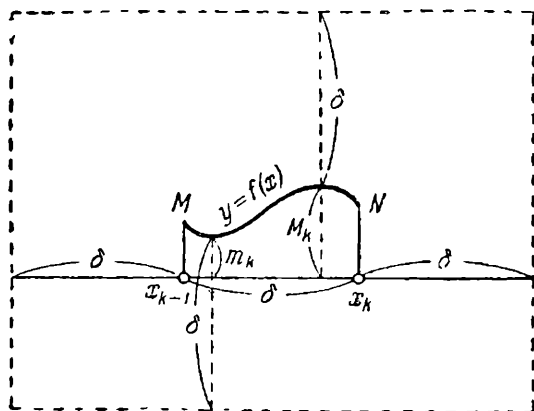


Fig. 72

sure we can select nets of straight lines in arbitrary directions; if we take these directions parallel to the sides of the given rectangle, we can see directly that the measure of the rectangle coincides with its area as determined by elementary methods. It therefore follows from theorem 2 that the same thing also holds for every figure composed of a finite number of rectangles. In chapter 12 we have used these figures to illustrate the "upper sums" $S(T)$ and the "lower sums" $s(T)$ of the given division T . Hence measure of a curvilinear trapezium (composed of the first of these figures and containing the second) is confined between $s(T)$ and $S(T)$ for every division T . And since the same also holds for the area of this curvilinear trapezium, the absolute value of the difference between measure and area does not exceed $S(T) - s(T)$; hence if the function $f(x)$ is continuous, it is equal to zero, since the difference $S(T) - s(T)$ can be made as small as we please by suitably choosing the division T .

Let us finally consider a curve expressed by the parametric equations

$$x = \varphi(t), \quad y = \psi(t), \quad (3)$$

where the functions φ and ψ are continuous and have continuous derivatives in an interval (t_0, t_1) of variation of the parameter t and assume that we have simultaneously $\varphi'(t) = \psi'(t) = 0$ at no point in this interval. Let t be an arbitrary point in the interval (t_0, t_1) and let $\varphi'(t) \neq 0$; because of continuity the function $\varphi'(t)$ retains its sign in a neighbourhood of the point t and therefore the function $x = \varphi(t)$ is continuous and monotone in an interval which surrounds

the point t ; but we know (§ 23) that in this case t is a single-valued function of x in that interval and therefore $y = \psi(t)$ is also a single-valued function of x in that interval of the curve. Hence every point in the interval $t_0 \leq t \leq t_1$ can be surrounded by an interval in which the curve (3) can be expressed by either of the two equations: $y = f_1(x)$, $x = f_2(y)$ (where the functions f_1 and f_2 are continuous in the respective intervals).

If we apply the theorem on finite coverage (lemma 2§ 18) to the set of these intervals we find that the curve (3) consists of a finite number of intervals, in each of which it can be represented by either of the two equations: $y = f_1(x)$ and $x = f_2(y)$ (where f_1 and f_2 are continuous). Theorem 6 thus leads to the following theorem:

Theorem 7. *Let the contour of the figure F be divisible into a finite number of intervals each of which can be expressed by equations of the form (3), where the functions $\varphi(t)$ and $\psi(t)$ have continuous derivatives in the corresponding intervals which do not vanish simultaneously. If this is so, then the figure F is measurable.*

§ 115. Volumes of cylindrical bodies

We have mentioned in chapter 12 that the development of integral calculus for functions of one variable was prompted by evaluation of areas of plane figures bounded by contours of arbitrary form; at the time we have used this example as an introduction to the concept of an integral. The calculation of volumes of bodies bounded by surfaces of arbitrary form plays an analogous role for functions of two variables. In this case too elementary geometry teaches us very little; apart from polyhedra (*i. e.* bodies bounded exclusively by planes) it only deals with volumes of bodies whose boundaries consist (sometimes in conjunction with parts of planes) of parts of spherical (ball-like), conical and cylindrical surfaces. We shall begin by giving a definition and developing a method for calculation of volumes of bodies bounded by surfaces which are in general arbitrary in shape.

As in evaluation of areas (*cf.* chapter 12) where we began by considering the general problem in relation to evaluation of areas of figures of specific shape (curvilinear trapeziums) we shall again concentrate our attention on calculation of volumes of bodies of specific form which we shall call "cylindrical" bodies. It can be readily seen that any body of a more or less simple form can be divided into several such cylindrical bodies and, therefore, by knowing to find

volumes of these cylindrical bodies, we shall have no difficulties in calculating volume of any body whose boundaries are not too complicated.

Let a measurable figure D (fig. 73) be given in a plane which we shall take as the coordinate plane XOY . Let a surface be given in space to extend over the figure D so that every straight line parallel to the OZ -axis intersects it at one point only; the equation of such a surface can be written in the form

$$z = f(x, y), \quad (1)$$

where we assume that the function $f(x, y)$ is positive and continuous in a rectangle which surrounds the figure D . Let us now draw a perpendicular from every point of the contour of the figure D to the XOY plane and continue it to

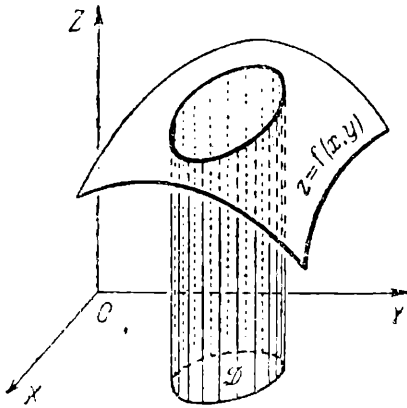


Fig. 73.

intersect the surface (1). The set of these straight lines represents the surface of the cylindrical figure whose generating lines are parallel to the OZ -axis. The resulting body which is bounded from below by the figure D , from sides by the cylindrical surface which we have just described and from above by the surface (1) is a cylindrical body; the definition and development of a method for calculation of volume of such a body is the object of this paragraph. It is evident that this problem is analogous to calculation of area of a curvilinear trapezium. The complication in this case is mainly due to the fact that a curvilinear trapezium is always bounded from below by a section of a straight line whereas the lower boundary of a cylindrical body can consist of the measurable figure D which can be arbitrary in form. However, this complication is due to the nature of the two-dimensional continuum: on a straight line (*i.e.* in a one-dimensional continuum) only a single form of a measurable figure is possible, *viz.* an interval whereas in a plane, even if we restrict ourselves to the simplest figures, we immediately meet an infinite variety of shapes. For this reason it is desirable to consider right from the beginning an arbitrary measurable figure as the "region of integration".

The same difference again appears as soon as we try to solve our problem *i.e.* when we try to define volume of a cylindrical body in analogy to area of a curvilinear trapezium. In the latter case we begin by dividing the interval (a, b) , which is the lower boundary

of our trapezium, into arbitrary parts which we called "cells" of the given division. Therefore in this case we naturally begin by arbitrarily dividing the figure D into figures which we shall also call cells. But which is the most suitable form for these cells? We now evidently have an indefinitely wider choice than before. Let us however, note that in the former case we found it useful not to restrict our choice of cells so that our arguments could apply to all possible divisions of the given interval. For the same reason we shall in this case also try not to restrict our choice in any way. We shall naturally demand that each cell should have a definite area, *i.e.* it should be a measurable figure; we shall assume further that no two cells should have interior points in common; in all other respects the dimensions, shapes and mutual positions of cells can be arbitrary. Let us denote the cells of the given division T of the figure D performed in any given order by $\Delta_1, \Delta_2, \dots, \Delta_n$; let Δ_k denote measure of the cell of the same description.

Let us now choose an arbitrary point in every cell $\Delta_k (\xi_k, \eta_k)$. The product

$$f(\xi_k, \eta_k) \Delta_k \quad (2)$$

expresses volume of a right cylinder with the base Δ_k and height $f(\xi_k, \eta_k)$. If the cell Δ_k is small, then it follows from continuity of the function $f(x, y)$ that the values of this function at different points in the cell Δ_k will differ very slightly from one another and therefore very little from the quantity $f(\xi_k, \eta_k)$. If we now cut, subjectively, a narrow cylindrical column situated above the cell Δ_k from our cylindrical body, then volume of this column will evidently differ very little from the volume of a straight cylinder with base Δ_k and height $f(\xi_k, \eta_k)$, *i.e.* from the quantity (2). And if all the cells are small, then volume of the whole cylindrical body which is equal to sum of volumes of all such columns will only differ very little from the sum

$$\sum_{k=1}^n f(\xi_k, \eta_k) \Delta_k. \quad (3)$$

We must emphasize once again that the division T of the figure D remains arbitrary (provided the cells Δ_k are measurable and sufficiently small); the choice of the points (ξ_k, η_k) in individual cells is also arbitrary.

It is evident that in future we must try to make the division T progressively "fine". But what does it mean? Let us note that in

the former case we have evaluated "fineness" of the division T by smallness of the quantity $l(T)$ in the greatest of its cells. We must naturally act similarly in this case. But how can we evaluate the dimensions of the cell? It can readily be seen that measure of the cell is not suitable for this purpose; in fact, in this case we try to make the cells small in order that any two points of a given cell should lie close to one another; however, the small dimensions of a cell do not guarantee us this property (the cell may have the shape of an elongated rectangle). Let us note that we have agreed to call the diameter of a plane figure as the upper bound of the common distances of all possible pairs of its points. Therefore smallness of diameter of the cell Δ_k is, in fact, necessary and sufficient in order that the common distance of two arbitrary points in it should be small. If we denote by $d(T)$ the greatest of the diameters of the cells of the given division T , then "fineness" of this division can conveniently be measured and evaluated by smallness of the quantity $d(T)$. We shall naturally say that the variable division T becomes "infinitely fine" if and only if $d(T) \rightarrow 0$.

All that follows is quite clear from its analogy to the definition of area of a curvilinear trapezium. If, when the division T becomes infinitely fine and choice of the points (ξ_k, η_k) in individual cells is arbitrary, the sum (3) tends to a limit V independent of the division T and choice of the points (ξ_k, η_k) in individual cells of the given division, we shall call this limit V *volume* of the given cylindrical body and write:

$$V = \lim_{d(T) \rightarrow 0} \sum_{k=1}^n f(\xi_k, \eta_k) \Delta_k. \quad (4)$$

It is clear from above that the exact meaning of this notation (in agreement with the accepted general concept of limiting process § 15) involves the following fact: a $\delta > 0$ can be found for arbitrarily small $\varepsilon > 0$ such that for every division T , where $d(T) < \delta$, and for every choice of the points (ξ_k, η_k) the following inequality holds for cells of the given division:

$$\left| \sum_{k=1}^n f(\xi_k, \eta_k) \Delta_k - V \right| < \varepsilon.$$

We have thus established the definition of volume of a cylindrical body. The apparatus for calculation of this volume follows

directly from the above definition, which is fully constructive ; however, in practice, this method because of its complexity gives us even less than the method obtained earlier in connection with the definition of area of a curvilinear trapezium. For this reason we must say that we have still not developed a suitable practical method for evaluation of volumes of cylindrical bodies.

From above we can draw no conclusions as to the conditions under which the limit involved in the definition of volume of a cylindrical body exists and is independent of the elements of construction. We must therefore still consider all these problems.

§ 116. Double integral

As in the one-dimensional case (chapter 14) we can now enumerate many geometrical and physical problems whose solution involves evaluation of limits of the form (4) § 115 (masses of thin heterogeneous plates, centres of gravity and moments of inertia of such plates, etc*). We must therefore study the general properties of these limits and find practical methods for their evaluation.

We shall naturally begin by agreeing on terminology and system of notation. If the given function $f(x, y)$ and the given measurable figure D have a limit of the form (4) § 115, then this limit is called *double integral of the function $f(x, y)$ in the "region of integration" D* and denoted by

$$\iint_D f(x, y) d\sigma \quad \text{or} \quad \iint_D f(x, y) dx dy.$$

In the first formula the symbol $d\sigma$ ('element of area') should remind us of origin of the integral obtained from "integral sums" as a result of the limiting process

$$\sum_{k=1}^n f(\xi_k, \eta_k) \Delta \sigma_k.$$

We know that the shape of cells in this process is arbitrary ; it is mostly convenient to have cells in the forms of rectangles with sides parallel to the axes of coordinates ; the integral sums then have the form :

$$\sum_{k=1}^n f(\xi_k, \eta_k) \Delta x \Delta y,$$

*) We shall consider some of these problems in § 120.

and the second formula of the double integral given above, which we shall mainly use in future, should remind us of origin of the integral. It is obvious that the meaning of both notations is the same.

If an integral exists, the function $f(x, y)$ is said to be *integrable* in the region D . Nothing is required of the function $f(x, y)$ except that it should be defined at every point of the region of integration D and be bounded in that region. Thus the function must not even be continuous; it may also assume negative values (in these more general cases the simple geometrical interpretation of double integral given in § 115 does not apply).

As in the one-dimensional case, the concepts of the upper and lower sums which are constructed in full analogy to the one-dimensional case (§ 47) are of great assistance in constructing of the theory of double integrals.

Let M and m denote respectively the upper and lower bounds of the function $f(x, y)$ in the region D . Let us assume that we have performed an arbitrary division T of this region into (measurable) cells $\Delta_1, \Delta_2, \dots, \Delta_n$ and let M_k and m_k denote the upper and lower bounds of the function $f(x, y)$ in the cell Δ_k respectively. We can thus construct sums

$$S_T = \sum_{k=1}^n M_k \Delta_k, \quad s_T = \sum_{k=1}^n m_k \Delta_k,$$

which are uniquely defined by the chosen division T called the *upper and lower sums* of this division. These sums possess all properties of the upper and lower sums in the one-dimensional case (§ 47, properties 1° – 4°); all previous proofs remain valid except that the cells Δ_k are now measurable plane figures. Instead of the length $b - a$ of the interval (a, b) we must now use everywhere the area D of the region of that name. In the proof 3° we must note that the cells Δ_{ki} are measurable figures in accordance with theorem 3 § 114.

As in the case of a one-dimensional integral we shall agree to call the quantity $\omega_k = M_k - m_k$ variation of the function $f(x, y)$ in the cell Δ_k of the division T . It can again be shown that the simple relation

$$S_T - s_T = \sum_{k=1}^n \omega_k \Delta_k \rightarrow 0 \quad [d(T) \rightarrow 0] \quad (1)$$

is a necessary and sufficient condition for integrability of the function $f(x, y)$ in the region D . The proof is exactly the same as in the one-dimensional case (§ 48).

The most important consequence of this criterion of integrability is, as in the one-dimensional case, integrability of all continuous functions. The function $f(x, y)$ continuous in the bounded closed region D is, as we know (§ 88), *uniformly continuous* in that region. Therefore provided $d(T)$ is sufficiently small, the variation ω_k of the function $f(x, y)$ in any cell Δ_k of the division T will be less than an arbitrary preassigned positive number ε which can be as small as we please; but this implies that, provided $d(T)$ is sufficiently small,

$$\sum_{k=1}^n \omega_k \Delta_k < \varepsilon \sum_{k=1}^n \Delta_k = \varepsilon D;$$

since $\varepsilon > 0$ is arbitrarily small, the relation (1) holds and the function $f(x, y)$ is integrable in the region D .

We have seen in the case of simple (single) integrals that a finite number of points of discontinuity of a bounded function does not effect its integrability (§ 48, theorem 4). We can similarly show that the bounded function $f(x, y)$ is integrable in every region D in which all its points of discontinuity are situated *on a finite number of lines of relatively simple type*. However, we shall not do so here.

Double integrals possess several simple properties which are completely analogous to those of simple integrals; the proof of these properties is usually quite simple and follows the same lines as the corresponding proofs for simple integrals. It is therefore sufficient only to enumerate the more important properties and leave the reader to provide the proof *).

1°. If the functions $f_1(x, y)$ and $f_2(x, y)$ are integrable in the region D , then their sum is also integrable in that region and

$$\begin{aligned} \iint_D [f_1(x, y) + f_2(x, y)] dx dy &= \\ &= \iint_D f_1(x, y) dx dy + \iint_D f_2(x, y) dx dy. \end{aligned}$$

*) Let us note that in everything that follows any measurable plane figure can serve as the region of integration.

2°. If k is an arbitrary constant and the function $f(x, y)$ is integrable in the region D , then the function $k f(x, y)$ is also integrable in that region and

$$\int \int_D k f(x, y) dx dy = k \int \int_D f(x, y) dx dy.$$

3°. If the function $f(x, y)$ is integrable in each of the regions D_1 and D_2 , then it is also integrable in the region $D_1 + D_2$; if, at the same time, the regions D_1 and D_2 have no common interior points, then

$$\int \int_{D_1 + D_2} f(x, y) dx dy = \int \int_{D_1} f(x, y) dx dy + \int \int_{D_2} f(x, y) dx dy.$$

4°. If the functions $f_1(x, y)$ and $f_2(x, y)$ are integrable in the region D and if at every point of that region $f_1 \leq f_2$, then

$$\int \int_D f_1(x, y) dx dy \leq \int \int_D f_2(x, y) dx dy.$$

5°. If the function $f(x, y)$ is integrable in the region D , then the function $|f(x, y)|$ is also integrable in that region and

$$\left| \int \int_D f(x, y) dx dy \right| \leq \int \int_D |f(x, y)| dx dy.$$

6°. If the function $f(x, y)$ is integrable in the region D and if at every point of that region $m \leq f(x, y) \leq M$, then

$$m D^* \leq \int \int_D f(x, y) dx dy \leq M D^*,$$

where D^* denotes measure of the region D .

7°. (Mean value theorem). If the function $f(x, y)$ is continuous in the closed region D and this region is "connected", i.e. any two points of this region can be joined by an open polygon which wholly belongs to the region D , then a point (ξ, η) can be found in this region such that

$$\int \int_D f(x, y) dx dy = f(\xi, \eta) D^*.$$

In order to prove this we must join two points at which the function $f(x, y)$ assumes its greatest value M and its smallest value m by an open polygon lying entirely in the region D . We can regard f as a continuous function of an arbitrary suitably chosen parameter λ along that polygon (for example, λ can be defined as the length of an interval of the open polygon from the origin to the given point). We can then apply theorem 3 § 23 to this continuous function, whose values at the end points of the given section of variation of the parameter are respectively equal to M and m ; since it follows from 6° that

$$m \leq \frac{1}{D^*} \iint_D f(x, y) \, dx \, dy \leq M,$$

therefore on the drawn open polygon (and hence also in the region D) a point (ξ, η) can be found at which

$$f(\xi, \eta) = \frac{1}{D^*} \iint_D f(x, y) \, dx \, dy,$$

which was to be proved.

We must finally draw attention to an important corollary of theorem 5 § 114 which we shall find useful on many future occasions. Let the function $f(x, y)$ be integrable in the region D ; its integral is defined as limit of sums of the form

$$\sum_k f(\xi_k, \eta_k) \Delta_k$$

in all the cells Δ_k of the given division T of the region D , provided the greatest diameter of the cells tends to zero. Among the cells Δ_k we distinguish inner and boundary cells; let Σ_1 denote summation in all inner cells and Σ_2 summation in all boundary cells of the given division so that

$$\sum_k f(\xi_k, \eta_k) \Delta_k = \sum_1 f(\xi_k, \eta_k) \Delta_k + \sum_2 f(\xi_k, \eta_k) \Delta_k$$

and

$$\sum_k \Delta_k = \sum_1 \Delta_k + \sum_2 \Delta_k.$$

If μ denotes the upper bound of the function $|f(x, y)|$ in the region D , then

$$\left| \sum_2 f(\xi_k, \eta_k) \Delta_k \right| \leq \mu \sum_2 \Delta_k = \mu \left(\sum_k \Delta_k - \sum_1 \Delta_k \right).$$

But theorem 5 § 114 maintains that if the division becomes indefinitely fine

$$\sum_1 \Delta_k \rightarrow \sum_k \Delta_k$$

(the above sum is obviously the measure of the region D); therefore if $d(T) \rightarrow 0$,

$$\sum_2 f(\xi_k, \eta_k) \Delta_k \rightarrow 0,$$

and consequently

$$\iint_D f(x, y) dx dy = \lim \sum_k f(\xi_k, \eta_k) \Delta_k = \lim \sum_1 f(\xi_k, \eta_k) \Delta_k.$$

Theorem. *If the function $f(x, y)$ is integrable in the region D , then*

$$\iint_D f(x, y) dx dy = \lim_{d(T) \rightarrow 0} \sum_1 f(\xi_k, \eta_k) \Delta_k,$$

where the sum \sum_1 only includes inner cells of the division T .

Note. The proved theorem remains valid if, apart from all inner cells, we include some (any we please) boundary cells in the sum \sum_1 (the sum \sum_2 then contains the remaining boundary cells). The proof remains unchanged.

§ 117. Evaluation of double integrals by means of two simple integrations

We have already mentioned above that our definition of a double integral gives a method for its evaluation but this method is rather limited because of its bulkiness and unsuitability for practical use. We must therefore develop a suitable method for evaluation of the given double integral. In most cases evaluation is achieved by reducing the double integral into two successive simple (*i.e.* single) integrals. We shall establish this general method in this paragraph.

Let the closed region (measurable figure) D where the given integral is defined has such a form that any straight line parallel to

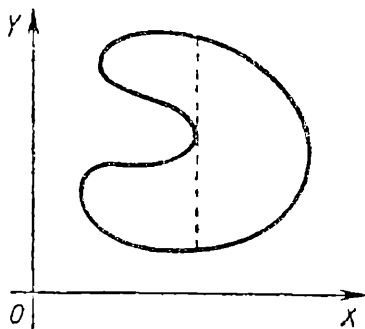


Fig. 74.

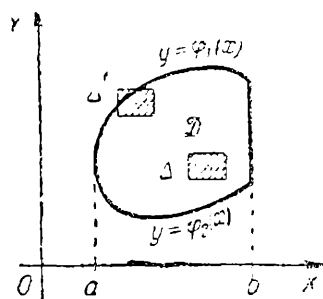


Fig. 75.

one of the coordinate axes intersects its contour only at two points (fig. 74; this condition excludes all regions similar to that represented in fig. 75; however, such a region, provided it is sufficiently simple, can always be divided into parts of the required form as shown by dotted lines in fig. 75); only lines at the extreme left and the extreme right permit this exception; each of these can share a whole section of common points with the boundary of the region D (as, for example, the line at the extreme right in fig. 74).

We shall assume that the function $f(x, y)$ is continuous in the region D . In that case the integral

$$I = \iint_D f(x, y) \, dx \, dy$$

is known to exist and (as a result of the last theorem of the previous paragraph, see note) it is equal to the limit of sums of the form for $d(T) \rightarrow 0$,

$$\sum_1 f(\xi_k, \eta_k) \Delta_k$$

(where ξ_k, η_k are coordinates of a point arbitrarily chosen in the cell Δ_k of the division T and the sum includes all inner cells and any number of boundary cells of this division); this limit is independent of the division T and choice of the points (ξ_k, η_k) in cells of these divisions. Therefore in order to obtain the reduction necessary for evaluation of the integral I and subsequent evaluation of two simple integrals we can choose the division T and the points (ξ_k, η_k) in the cells of these divisions in a way most convenient for our purpose if only $d(T) \rightarrow 0$.

Let $\varepsilon > 0$ be given arbitrarily; in that case provided $\delta > 0$ is sufficiently small, we have

$$\left| I - \sum_1 f(\xi_k, \eta_k) \Delta_k \right| < \varepsilon \quad (1)$$

for every division T for which $d(T) < \delta$ (and for every choice of the points (ξ_k, η_k) in the cells of this division). Let us now choose one such division in a definite manner. Let us write respectively $y = \varphi_1(x)$ and $y = \varphi_2(x)$ for the equations of the upper and lower parts of the contour of the region D (fig. 74) and assume that these two functions are continuous in the interval (a, b) between the terminal abscissae of points of the region D . It then follows from the theorem on uniform convergence that an $h_0 > 0$ can be found such that when $a \leq x' < x'' \leq b$ and $|x' - x''| < h_0$, we have:

$$|\varphi_1(x') - \varphi_1(x'')| < \frac{\delta}{2}, \quad |\varphi_2(x') - \varphi_2(x'')| < \frac{\delta}{2}. \quad (2)$$

Let us now divide the interval (a, b) by means of the following points of division:

$$a = x_0, x_1, \dots, x_n = b$$

into a large number of equal parts so that

$$x_i - x_{i-1} = h < h_0 \quad (1 \leq i \leq n)$$

and also $h < \delta/2$. Let us draw a straight line parallel to the OT -axis through each point of division. The set of these straight lines divides the region D into vertical strips (fig. 76). A further choice of cells of division will be made separately in each strip.

Fig. 77 represents one of these vertical strips confined between the straight lines $x = x_i$ and $x = x_{i+1}$. Let M_1 and m_1 denote respectively the greatest and least values of the function $\varphi_1(x)$ in the given strip while M_2 and m_2 have similar meanings for the function $\varphi_2(x)$. Since $h < h_0$, it follows from the inequalities (2) that

$$M_1 - m_1 < \frac{\delta}{2}, \quad M_2 - m_2 < \frac{\delta}{2}. \quad (3)$$

Let us now draw in our strip (Fig. 77) straight lines parallel to the OX -axis at the heights m_2, M_2, m_1, M_1 (let us assume for the sake of simplicity that $M_2 < m_1$) and divide the interval $M_2 \leq y \leq m_1$ of the OT -axis into equal parts and draw straight lines parallel to the OX -axis through each point of division. These straight lines divide

the part of the region D confined in the given strip into parts which we shall regard as cells of our division T . We have not yet defined

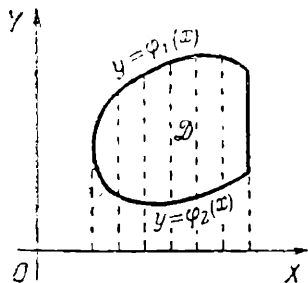


Fig. 76

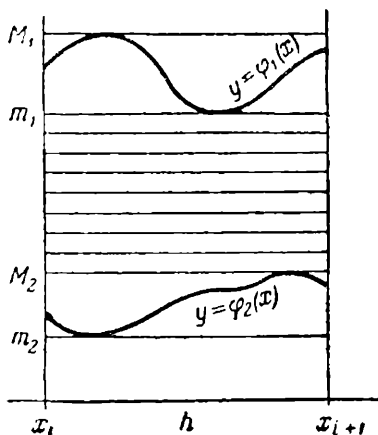


Fig. 77

the number of equal parts into which the interval $M_2 \leq y \leq m_1$ should be divided; we shall do so now.

In each cell Δ_k constructed in the given strip as described above let us select $\xi_k = x_i$ (which will be the same for all cells of the given strip) whereas the numbers η_k can be chosen arbitrarily. The lowest cell of the given strip is contained in a rectangle whose dimensions are $h < \delta/2$ and $M_2 - m_2 < \delta/2$ and therefore its diameter does not exceed $\sqrt{h^2 + (M_2 - m_2)^2} < \delta/\sqrt{2} < \delta$. The same evidently also holds for the highest cell. All remaining cells of the given strip are rectangles and the diameter of each is less than δ , provided we divide the interval $M_2 \leq y \leq m_1$ into a sufficiently large number of (equal) parts. From these rectangles the lowest and highest are evidently boundary cells whereas the remaining ones are inner cells of our division.

Part of the integral sum

$$\sum_k f(\xi_k, \eta_k) \Delta_k,$$

which includes the rectangular cells of our chosen strip can be written in the form

$$\sum_{l=0}^{m-1} f(x_i, y_l) h h',$$

where h' is the distance between two adjacent points of division of the interval (M_2, m_1) and y_l is an arbitrary number confined between two such adjacent points of division.

But the sum

$$\sum_{l=0}^{m-1} f(x_i, y_l) h'$$

for $h' \rightarrow 0$ has the integral

$$\int_{M_2}^{m_1} f(x_i, y) dy.$$

as its limit. We therefore have, provided h' is sufficiently small :

$$\left| \sum_{l=0}^{m-1} f(x_i, y_l) h' - h \int_{M_2}^{m_1} f(x_i, y) dy \right| < h\varepsilon. \quad (4)$$

But since

$$m_1 \leq \varphi_1(x_i) \leq M_1, \quad m_2 \leq \varphi_2(x_i) \leq M_2,$$

therefore it follows from the inequalities (3)

$$\varphi_1(x_i) - m_1 < \frac{\delta}{2}, \quad M_2 - \varphi_2(x_i) < \frac{\delta}{2},$$

and hence

$$\begin{aligned} & \left| \int_{\varphi_2(x_i)}^{\varphi_1(x_i)} f(x_i, y) dy - \int_{M_2}^{m_1} f(x_i, y) dy \right| = \\ & = \left| \int_{\varphi_2(x_i)}^{M_2} f(x_i, y) dy + \int_{m_1}^{\varphi_1(x_i)} f(x_i, y) dy \right| < 2\mu \frac{\delta}{2} = \mu\delta, \end{aligned}$$

where μ is the upper bound of the function $|f(x, y)|$ in the region D . It therefore follows from (4)

$$\left| \sum_{l=0}^{m-1} f(x_i, y_l) h' - h \int_{\varphi_2(x_i)}^{\varphi_1(x_i)} f(x_i, y) dy \right| < h\varepsilon + h\mu\delta < 2h\varepsilon,$$

since there is no reason why we should not choose $\delta < \varepsilon/\mu$.

We have so far considered part of the integral sum which embraced the rectangular cells in one vertical strip chosen by us. Summing the result over all such vertical strips we obtain :

$$\sum_1 f(\xi_k, \eta_k) \Delta_k - h \sum_{i=0}^{n-1} \int_{\varphi_2(x_i)}^{\varphi_1(x_i)} f(x_i, y) dy \Big| < 2h\varepsilon n = 2\varepsilon(b-a),$$

where the sum Σ_1 includes all inner and some boundary cells of the given division.

Let us now assume that

$$\int_{\varphi_2(x)}^{\varphi_1(x)} f(x, y) dy = F(x) \quad (a \leq x \leq b),$$

so that $F(x)$ is a continuous function of x in the interval (a, b) (see theorem 4 § 109). The last inequality can then be written in the form

$$\sum_1 f(\xi_k, \eta_k) \Delta_k - \sum_{i=0}^{n-1} F(x_i) (x_{i+1} - x_i) \Big| < 2\varepsilon(b-a).$$

But the sum

$$\sum_{i=0}^{n-1} F(x_i) (x_{i+1} - x_i) \rightarrow \int_a^b F(x) dx$$

as $h \rightarrow 0$; we can therefore right from the beginning take h so small that this sum should differ from the integral

$$\int_a^b F(x) dx = \int_a^b \left\{ \int_{\varphi_2(x)}^{\varphi_1(x)} f(x, y) dy \right\} dx$$

by less than ε . The last inequality gives us :

$$\left| \sum_1 f(\xi_k, \eta_k) \Delta_k - \int_a^b F(x) dx \right| < \varepsilon [1 + 2(b-a)]. \quad (5)$$

Finally we see that all cells of the described division T have diameters less than δ and therefore the inequality (1) holds. But it follows from (1) and (5)

$$\left| I - \int_a^b F(x) dx \right| < 2\varepsilon(1 + b - a).$$

Since ε is as small as we please and the left-hand side of the last inequality does not depend on ε , it is equal to zero and we obtain :

$$\int_D \int f(x, y) dx dy = \int_a^b \left\{ \int_{\varphi_2(x)}^{\varphi_1(x)} f(x, y) dy \right\} dx. \quad (6)$$

This is the result at which we were aiming. We thus see that a double integral of a continuous function can be evaluated by means of two successive simple (single) integrations. On the right-hand side of formula (6) the inner integral is constructed in the way described in § 109 : the variable x plays the part of a parameter in this integral, *i.e.*, it retains a constant value during integration ; not only the integrand but also both limits of integration depend on this parameter x . Hence the inner integral as a whole is a continuous function of the variable x and must therefore be integrated with respect to x from a to b .

It is obvious that the coordinates x and y are completely equivalent in all these arguments, and we can, if for some reason it is more convenient, integrate first with respect to x and then with respect to y ; for this purpose we must select points on the contour D at which the ordinate assumes its least value $y = c$ and its greatest value $y = d$; these two points divide the contour into two parts — the *right-hand side* whose equation is $x = \psi_1(y)$ and the *left-hand side* with the equation $x = \psi_2(y)$. We thus obtain as above :

$$\int_D \int f(x, y) dx dy = \int_c^d \left\{ \int_{\psi_2(y)}^{\psi_1(y)} f(x, y) dx \right\} dy. \quad (7)$$

Evidently it is not always easy to perform the two necessary simple integrations. However, in principle we have converted evaluation of a double integral into a problem which we have already studied and we can therefore regard our problem solved. Further, having at our disposal two equivalent formulae (6) and (7) we can, of

course, in each case choose the formula which is most convenient for our purpose; this choice depends on the nature of the function $f(x, y)$ and the form of the region D .

Example. The trihedral prism whose generating lines are parallel to the OZ -axis, has its base in the XOY plane in the form of a triangle with vertices at the points $A(0, 1)$, $B(1, 0)$, $C(-1, 0)$. Find the volume of part of this prism confined between the plane XOY and the paraboloid of rotation $z = x^2 + y^2$ (fig. 78).

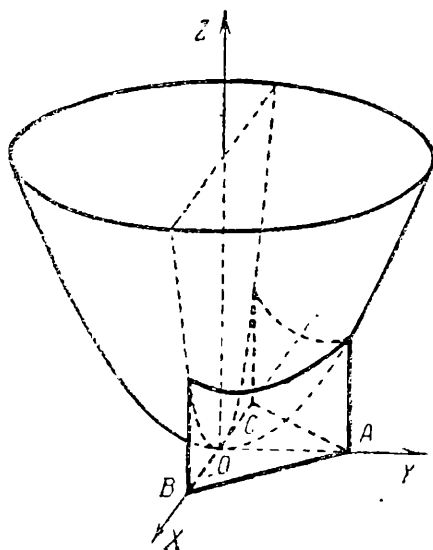


Fig. 78

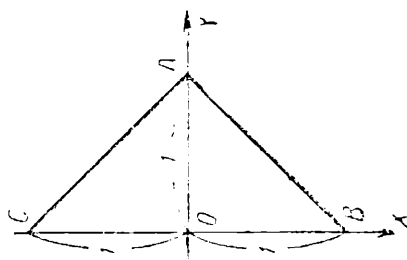


Fig. 79

We evidently have here a cylindrical body of the kind considered in § 115. Here the base ABC of the prism is the region D as shown in fig. 79. The equations of the straight lines AC and AB are respectively

$$x = y - 1, \quad x = 1 - y;$$

we therefore obtain the following expression for the required volume

$$V = \iint_D (x^2 + y^2) dx dy = \int_0^1 \left\{ \int_{y-1}^{1-y} (x^2 + y^2) dx \right\} dy.$$

Here the inner integral is

$$\int_{y-1}^{1-y} (x^2 + y^2) dx = \left[y^2 x + \frac{x^3}{3} \right]_{y-1}^{1-y} = 2y^2(1-y) + \frac{2}{3}(1-y)^3,$$

and we obtain :

$$V = \int_0^1 \left\{ 2y^2(1-y) + \frac{2}{3}(1-y)^3 \right\} dy.$$

Further calculations present no difficulties.

For further useful exercises cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 7-15, 17, 23.

§ 118. Substitution of variables in double integrals

We already know the significance of transformation of the variable of integration in evaluation of simple (single) integrals (we only need to remind you of the rationalisation of integrals of irrational and transcendental functions carried out in a large number of examples in chapter 17 where whole classes of functional dependencies were involved). Owing to the great freedom of choice of transformations we are able to replace in most cases the integrand by another simpler expression which can be integrated more readily.

We must now develop a similar method for double integrals. We shall see that here a simple method for transformation of the variables of integration is also possible and this frequently facilitates evaluation of the integral.

Let us assume that we are given the double integral

$$\iint_D f(x, y) dx dy, \quad (1)$$

where the function $f(x, y)$ is continuous in the closed region D .

Let us consider, apart from the xy -plane, another uv -plane and a region D' in that plane. Let the transformation of the variables

$$x = x(u, v), \quad y = y(u, v), \quad (2)$$

where the functions $x(u, v)$, $y(u, v)$ are defined continuous and have continuous partial derivatives in the region D' , transform this region into the region D in the xy -plane (as considered in detail in § 105). We shall assume that the transformation of the region D' into the region D , as effected by the relations (2), is 1—1 i.e. that every point (x, y) in the region D can only be transformed into a *single* point (u, v) in the

region D' ; that point (u, v) is given by the "reciprocal" transformation

$$u = u(x, y), \quad v = v(x, y), \quad (3)$$

in which we also assume that the functions $u(x, y)$, $v(x, y)$ are continuous and have continuous partial derivatives in the region D . Since in this case (cf. § 105)

$$\frac{D(u, v)}{D(x, y)} \cdot \frac{D(x, y)}{D(u, v)} = 1,$$

therefore both determinants are non-zero in the corresponding regions. We want to express the integral (1) over the region D in the xy -plane by another double integral over the region D' in the uv -plane.

For this purpose we shall consider the division T of the region D effected by two families of straight lines parallel to the OX and OY axes; let the distance h between two adjacent lines be the same for both families of lines so that the inner cells of the division T are squares with side h . This division corresponds to a definite division T' of the region D' whose cells will, in general, have curvilinear contours. Let us take an arbitrary inner cell (square) Δ_k of the division T in the region D with vertices at the points (x_k, y_k) , $(x_k + h, y_k)$, $(x_k, y_k + h)$, $(x_k + h, y_k + h)$; this cell will, in accordance with the transformation (3), correspond to another cell Δ'_k of the division T' in the region D' . Let us first show that each cell is a measurable figure.

The contour of the cell Δ'_k can be divided into four parts in relation to the four sides of the square Δ_k^*). Let us consider one of these four parts, for example the part corresponding to the side $x_k \leq x \leq x_k + h$, $y = y_k$ of the square Δ_k . This side can evidently be expressed by the parametric equations

$$u = u(x, y_k), \quad v = v(x, y_k) \quad (x_k \leq x \leq x_k + h),$$

where the derivatives $\partial u / \partial x$ and $\partial v / \partial x$ are continuous in the interval $(x_k, x_k + h)$ and cannot vanish simultaneously, since otherwise the determinant $D(u, v) / D(x, y)$ would also vanish, as we see, is impos-

*Here and latter in order to give a strict theoretical basis to our arguments we should show that in a 1—1 and mutually continuous transformation of one region into another the boundary of one region is transformed into that of the other region. This proposition is true but its proof lies beyond the scope of our course.

sible. Since the same also holds for each of the three remaining sides of the contour of the cell Δ_k , therefore, in accordance with theorem 7 § 144 this cell is a measurable figure.

In § 105 we have seen that the ratio Δ'_k/Δ_k of the areas of two cells has as its limit the absolute value of the following determinant for $h \rightarrow 0$

$$\frac{D(u, v)}{D(x, y)}$$

at the point (x_k, y_k) ; thus for $h \rightarrow 0$

$$\Delta'_k = \left| \frac{D(u, v)}{D(x, y)} \right| \Delta_k + o(\Delta_k).$$

Therefore, conversely

$$\Delta_k = \left| \frac{D(x, y)}{D(u, v)} \right| \Delta'_k + \alpha_k \Delta_k, \quad (4)$$

where $\alpha_k \rightarrow 0$ as $h \rightarrow 0$ and Ostrogradskij's determinant should be taken at the point $[u_k = u(x_k, y_k), v_k = v(x_k, y_k)]$ in the cell Δ'_k . Let us multiply the equation (4) by $f(x_k, y_k)$ and sum it over all cells of the region D :

$$\begin{aligned} \sum_1 f(x_k, y_k) \Delta_k &= \\ &= \sum_1 f(x_k, y_k) \left| \frac{D(x, y)}{D(u, v)} \right| \Delta'_k + \sum_1 f(x_k, y_k) \alpha_k \Delta_k. \end{aligned} \quad (5)$$

Let us now decrease the number h indefinitely and hence also the greatest diameter $h\sqrt{2}$ of the cells Δ_k of the division T . It follows from the final theorem of § 116 that the left-hand side of the equation will have the integral (1) as its limit. Let us investigate the right-hand side and begin by considering the second term. We have seen in § 105 that the numbers α_k tend to zero *uniformly* with respect to the position of the square Δ_k in the region D as $h \rightarrow 0$; therefore no matter how small $\varepsilon > 0$ and provided h is sufficiently small we shall have $|\alpha_k| < \varepsilon$ in all terms of the sum $\sum_1 f(x_k, y_k) \alpha_k \Delta_k$ and, therefore, provided h is sufficiently small:

$$|\sum_1 f(x_k, y_k) \alpha_k \Delta_k| < \varepsilon \sum_1 |f(x_k, y_k)| \Delta_k \leq \mu \varepsilon D,$$

where μ denotes the upper bound of the function $|f(x, y)|$ in the region D . Since ε is as small as we please, therefore as $h \rightarrow 0$

$$\lim \sum_1 f(x_k, y_k) \alpha_k \Delta_k = 0.$$

Let us finally consider the first sum on the right-hand side of the equation (5). Assuming, for the sake of brevity, that

$$\frac{D(x, y)}{D(u, v)} = \mathcal{J}(u, v)$$

and noting that

$$x_k = x(u_k, v_k), \quad y_k = y(u_k, v_k),$$

we can write this sum in the form

$$\sum_1 f[x(u_k, v_k), y(u_k, v_k)] |\mathcal{J}(u_k, v_k)| \Delta'_k, \quad (6)$$

where the sum extends over all cells Δ'_k in the region D' which correspond to inner cells Δ_k in the region D . But as a result of the transformation (3) the contour of the region D is transformed into the contour of the region D' and *vice versa* (see footnote, p. 585). It therefore follows that inner cells of one region should correspond to inner cells of the other region and the same holds for boundary cells. The sum (6), which is extended over the cells of the region D' corresponding to inner cells of the region D , must therefore be extended over all inner cells of the region D' . If $h \rightarrow 0$, the greatest diameter of these cells (as a result of uniform continuity of the functions $u(x, y)$, $v(x, y)$) tends to zero and, as a result of the last theorem of § 116, the sum (6) has as its limit the double integral *)

$$\iint_{D'} f[x(u, v), y(u, v)] |\mathcal{J}(u, v)| du dv. \quad (7)$$

Hence returning to the relation (5) and taking its limit for $h \rightarrow 0$ we obtain the formula

$$\iint_D f(x, y) dx dy = \iint_{D'} f[x(u, v), y(u, v)] |\mathcal{J}(u, v)| du dv, \quad (8)$$

where

$$\mathcal{J}(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

and where D' is a region in the uv -plane which is transformed into the region D in the xy -plane by the transformation $x = x(u, v)$, $y = y(u, v)$. This is the formula for the substitution of variables in

*) We do not prove here that the region D' , like the region D , is a measurable figure

double integrals which we were trying to obtain. We can see that it is analogous to the corresponding formula for the substitution of the variable in simple (single) integrals

$$\int_a^b f(x) dx = \int_{\alpha}^{\beta} f[\varphi(t)]\varphi'(t)dt,$$

where the interval $a \leq x \leq b$ is transformed into the interval $\alpha \leq t \leq \beta$ by the transformation $x = \varphi(t)$. The derivative $\varphi'(t)$ corresponds to the determinant \mathcal{J} in formula (8) or, more strictly, to its absolute value. As with simple integrals the main purpose of formula (8) is the transformation of integrals into forms more suitable for calculation; however, a new factor is involved here which has no parallel to single integrals: the use of formula (8) often involves not only simplification of the integrand but also effect the form of the region of integration; if the form of the region D' is simpler than that of the region D , then the integral will itself be greatly simplified and this simplification is so important that in order to achieve it, it is in some cases even desirable to complicate the integrand a little.

Example. The most frequent type of transformation of variables in a double integral is the transition from rectangular coordinates (x, y) to polar coordinates (r, φ) ; in the simplest cases the transformation formulae are as follows:

$$x = r \cos \varphi, \quad y = r \sin \varphi,$$

and therefore

$$\mathcal{J} = \frac{D(x, y)}{D(r, \varphi)} = \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} = r.$$

The general formula has the form

$$\iint_D f(x, y) dx dy = \iint_{D'} f[r \cos \varphi, r \sin \varphi] r dr d\varphi. \quad (9)$$

This formula is particularly convenient in case the region D is a circle with centre at origin:

$$x^2 + y^2 \leq a^2; \quad (10)$$

it is evident that in this case the region D' is a *rectangle* $0 \leq r \leq a$, $0 \leq \varphi \leq 2\pi$. It is obviously much simpler to integrate round a

rectangle than round a circle; the reduction to two simple integrations as considered in § 117 gives for a circle (of the variables x, y) the following limits of integration

$$\int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} ;$$

and for a rectangle (of the variables r, φ) it gives constant limits:

$$\int_0^a \int_0^{2\pi}$$

which in many problems considerably simplifies calculations *).

Let us assume, for example, that we want to evaluate the integral

$$I = \iint_D e^{-x^2-y^2} dx dy,$$

where D is the circle (10). Formula (9) gives:

$$I = \iint_{D'} e^{-r^2} r dr d\varphi,$$

where D' is the rectangle $0 \leq r \leq a, 0 \leq \varphi \leq 2\pi$. It therefore follows from the results of § 117 that

$$\begin{aligned} I &= \int_0^{2\pi} \left\{ \int_0^a r e^{-r^2} dr \right\} d\varphi = 2\pi \int_0^a r e^{-r^2} dr = \\ &= 2\pi \left(-\frac{1}{2} e^{-r^2} \right) \Big|_0^a = \pi (1 - e^{-a^2}). \end{aligned}$$

The reader should try to evaluate the same integral in rectangular coordinates when he will encounter considerable difficulties.

For further exercises *cf.* Problem Book by B. P. Demidovich, Section VIII, Nos. 34-36, 39, 40, 47, 49, 51, 53, 54, 86, 87.

*) The fact that Ostrogradskij's determinant vanishes at the centre of the circle does not effect the validity of formula (9) as can readily be shown.

§ 119. Triple integrals

In the last few paragraphs we had occasion to observe that the transition [from simple integrals to double integrals necessitated a detailed study of the region of integration, *i.e.* we had to use the theory of measurable figures, in spite of the close analogy existing between these two types of integrals. The construction of the theory described in § 114 required considerable effort; however, this effort is amply rewarded by the simplicity and accuracy of the subsequent construction of the theory of double integrals; it is even more important to note that § 114 can entirely and almost without modifications be extended to measurable sets in space of any number of dimensions and that, consequently, the theory evolved in this paragraph can serve as the basis for integrals of any multiplicity.

In order to construct the theory for measuring sets in a three dimensional space we must follow the definitions and arguments of § 114 step by step; the necessary changes evidently involve the replacement of nets of straight lines by nets of planes, *i.e.* parallelograms are replaced by parallelopipeds, circles by spheres, etc; however, the arguments of § 114 are essentially preserved without modifications.

The triple integral

$$\iiint_V f(x, y, z) dx dy dz,$$

where V is an arbitrary bounded measurable set in the three dimensional space and $f(x, y, z)$ a function defined in the region V , is said to be limit of the sum

$$\sum f(\xi_k, \eta_k, \zeta_k) \Delta_k,$$

which extends over all cells Δ_k of a division T of the region V , where (ξ_k, η_k, ζ_k) is an arbitrary point in the cell Δ_k ; this limit is taken on the assumption that the greatest diameter of the division T tends to zero and is independent of the effected division T and choice of the points (ξ_k, η_k, ζ_k) in the cells of these divisions.

§ 116 can be extended to triple integrals without alterations. The practical evaluation of a triple integral is usually carried out by replacing a triple integral by *three* successive simple (single) integrations as we have seen in § 117 for double integrals. Thus we obtain for regions of sufficiently simple form

$$\iiint_V f(x, y, z) dx dy dz = \int_a^b \left\{ \int_{\varphi_2(x)}^{\varphi_1(x)} \left[\int_{\psi_2(x, y)}^{\psi_1(x, y)} f(x, y, z) dz \right] dy \right\} dx. \quad (1)$$

The inner integral is here taken with respect to z while x and y are parameters; the limits of this integral represent respectively the upper and lower bounds of values of z in the region V for *given* x and y ; therefore the inner integral (in square brackets) is a function of x and y ; this function is subsequently integrated with respect to y where x is a parameter; the limits of this integration are the bounds $\varphi_1(x)$ and $\varphi_2(x)$ of y in the region V for the given value of x ; a function of x (placed in crooked brackets) is obtained as a result of this second integration; finally (the third integration) this function is integrated with respect to x , where the limits are the bounds of x over the whole region V . It is evident that the above order of integration can be replaced by any other order (with corresponding changes of limits of integration) and this freedom in the order of integration should be used in each case to simplify the whole sequence of operations.

If we denote the inner integral on the right-hand side of formula (1) by $\Phi(x, y)$, then the right-hand side as a whole can be written in the form

$$\int_a^b \left\{ \int_{\varphi_2(x)}^{\varphi_1(x)} \Phi(x, y) dy \right\} dx,$$

which, according to § 117, is equal to

$$\iint_D \Phi(x, y) dx dy,$$

where D is the projection of the region V on the XOY -plane. Formula (1) therefore gives us:

$$\iiint_V f(x, y, z) dx dy dz = \iint_D \left\{ \int_{\psi_2(x, y)}^{\psi_1(x, y)} f(x, y, z) dz \right\} dx dy. \quad (2)$$

We thus see that *triple integration can be replaced by the successive performance of one single and one double integration.*

In order to get used to the quick evaluation, from the given geometrical region of integration V of the limits of all three simple

integrals on the right hand side of formula (1) and also to the solution of the converse problem determination of the form of the region V from the given limits of the simple integrals, considerable practice is necessary. Therefore good manuals on integral calculus contain many examples of this kind which are characterised by the fact that their solution not only involves integrations of any kind but even the type of the integrand must necessarily be known.

Finally the formula for the transformation of variables, which is analogous to formula (8) § 118, also holds for triple integrals. If the transformation

$$u = u(x, y, z), \quad v = v(x, y, z), \quad w = w(x, y, z) \quad (3)$$

maps 1-1 the region V in the xyz -space into the region V' in the uvw -space, then, provided the usual requirements of continuity and the condition $\mathcal{J} \neq 0$ are preserved, we have;

$$\begin{aligned} \iiint_V f(x, y, z) \, dx \, dy \, dz &= \\ &= \iiint_{V'} f[x(u, v, w), y(u, v, w), z(u, v, w)] |\mathcal{J}| \, du \, dv \, dw, \end{aligned}$$

where

$$\mathcal{J} = \frac{D(x, y, z)}{D(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

is Ostrogradskij's determinant of the transformation

$$x = x(u, v, w), \quad y = y(u, v, w), \quad z = z(u, v, w),$$

which is inverse of the transformation (3). The deduction of this formula (which we shall not attempt here) is carried out in close analogy to the deduction of formula (8) § 118 if we see in advance that in the three-dimensional case (as in the two dimensional case) the absolute value of Ostrogradskij's determinant can be interpreted geometrically as a specific "coefficient of expansion" in the transformation of infinitely small bodies.

As in the two-dimensional case, one of the most frequent types of transformation of variables in a triple integral is the transition from rectangular to spherical coordinates :

$$\begin{aligned}x &= r \cos \varphi \cos \psi, \\y &= r \cos \varphi \sin \psi, \\z &= r \sin \varphi,\end{aligned}$$

where

$$0 \leq r < +\infty, \quad -\frac{\pi}{2} \leq \varphi \leq +\frac{\pi}{2}, \quad 0 \leq \psi \leq 2\pi.$$

We readily obtain :

$$\left| \frac{D(x, y, z)}{D(r, \varphi, \psi)} \right| = r^2 \cos \varphi,$$

and the transformation formula takes the form

$$\begin{aligned}\iiint_V f(x, y, z) \, dx \, dy \, dz &= \\&= \iiint f(r \cos \varphi \cos \psi, r \cos \varphi \sin \psi, r \sin \varphi) r^2 \cos \varphi \, dr \, d\varphi \, d\psi.\end{aligned}$$

This transformation is usually most convenient if the region of integration is initially the sphere

$$x^2 + y^2 + z^2 \leq a^2,$$

and if the region V' is the rectangular parallelepiped $0 \leq r \leq a$, $-\pi/2 \leq \varphi \leq \pi/2$, $0 \leq \psi \leq 2\pi$.

For exercises to § 119 cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 148—150, 158, 159.

§ 120. Applications

In this paragraph we shall briefly consider several applications of double and triple integrals to geometrical and statical problems.

1. Area of a surface. The area of a given part of a curved surface can only in a few isolated cases be solved with the help of elementary geometry. In general, this problem involves the methods of integral calculus. We shall see that the part played by a double integral in the solution of this problem is analogous to that played by the usual integral in finding length of an arc of a curve (§ 52).

Let us assume that we are given a part of curved surface which is intersected only at one point by a straight line drawn in a given direction which we shall take as the direction of the OZ -axis. This part of the given surface can be expressed by the following equation:

$$z = f(x, y), \quad (1)$$

where we assume that the function $f(x, y)$ is continuous and has continuous partial derivatives with respect to x and y . Let the projection of the given part of the surface (1) onto the XOY -plane be a measurable figure D .

We know (§ 99) that with the assumptions made with regard to the function $f(x, y)$, the surface (1) has a tangential and a normal plane at every point of the given part; if we denote by γ the acute angle between this normal and the OZ -axis, then

$$\cos \gamma = \frac{1}{\sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}}. \quad (2)$$

As usual in geometrical applications of mathematical analysis our problem involves the definition of area of a part of a curved surface after which we must evolve an apparatus for the evaluation of this area. For this purpose let us divide the region D arbitrarily by the division T into cells for which we shall only demand that each cell must be a measurable figure and no two cells should have any common interior points. Let us select an arbitrary point (ξ_k, η_k) in every cell Δ_k and draw a perpendicular from this point to the XOY -plane to intersect the surface (1) at the point $M_k(\xi_k, \eta_k, \zeta_k)$, where $\zeta_k = f(\xi_k, \eta_k)$. Let us draw a tangential plane to the surface (1) at the point M_k . In the neighbourhood of the point M_k the course of this tangential plane is close to the course of the surface (1) itself; hence this visual representation shows us directly that, provided the diameter of the cell Δ_k is very small, the measure s_k of the part of the surface (1) projected onto the cell Δ_k should be very close to the measure σ_k of the drawn tangential plane projected onto the same cell. Summing this approximate equality over all cells we conclude that, provided the diameters of the cells are small, we are justified in taking the measure (area) of the whole part of the surface (1) in which we are interested (this measure is equal to $\sum_k s_k$) as being close to the sum

$$\sum_k \sigma_k; \quad (3)$$

therefore if this sum tends to a definite limit when the diameter of the cells become infinitely small, we naturally take this limit as the measure (area) of the part of the surface (1) in which we are interested.

We shall now show that if the division T becomes infinitely fine, the limit of the sum (3) always exists irrespective of the performed division and choice of the points (ξ_k, η_k) in the cells Δ_k . Let us note the meaning of the quantity σ_k . We have chosen an arbitrary point (ξ_k, η_k) in the cell Δ_k and assumed that $f(\xi_k, \eta_k) = \zeta_k$; we have also drawn a tangential plane to the surface (1) at the point $M_k(\xi_k, \eta_k, \zeta_k)$; σ_k denotes the measure of that part of this surface whose projection into the XOY -plane is the cell Δ_k . Since the angle between these two planes is evidently equal to the angle γ as considered above, therefore, in accordance with the general law connecting the area of projected figure and that of the projection*) we must have:

$$\Delta_k = \sigma_k \cos \gamma,$$

where the angle γ is obviously taken for the point M_k ; therefore as a result of (2)

$$\sigma_k = \frac{\Delta_k}{\cos \gamma} = \sqrt{1 + f'^2_x(\xi_k, \eta_k) + f'^2_y(\xi_k, \eta_k)} \Delta_k$$

and; consequently

$$\sum_k \sigma_k = \sum_k \sqrt{1 + f'^2_x(\xi_k, \eta_k) + f'^2_y(\xi_k, \eta_k)} \Delta_k$$

But this sum has exactly the same form as the sum considered in § 116; in accordance with the assumptions made with regard to $\partial f / \partial x$ and $\partial f / \partial y$ in agreement with the results of § 116 it therefore follows that

*) "The area of a projection is equal to the area of the projected figure multiplied by the cosine of the angle between the two surfaces in question". This rule is proved in elementary geometry for figures whose areas can be determined within the scope of this science. We have applied this rule to a more general case. In fact, we make use of the following proposition: "if the projection of the given figure is measurable, then the figure itself is measurable and the measure of the projection is equal to the measure of the projected figure multiplied by the cosine of the angle between the two surfaces in question". This general rule follows directly from the above rule of elementary geometry but we cannot go into this question in greater detail.

the sum under consideration tends to the following integral as its limit when the division T becomes infinitely fine :

$$S = \iint_D \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} dx dy, \quad (4)$$

which, according to the accepted definition, must be regarded as the expression of the area S of the part of the surface (1) in which we are interested. All necessary requirements concerned with the independence of the limit obtained from the elements of construction (*i.e.* from the choice of the divisions T and the points (ξ_k, η_k) in the cells) follow directly from the general theory of double integrals (§ 116).

It is interesting to compare (4) with the formula

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx,$$

which expresses the length L of an arc of a plane curve $y = f(x)$ confined between $x = a$ and $x = b$ and note the close analogy between these formulae.

Example. Find the surface area of a sphere of radius a . Let us place the origin of a rectangular system of coordinates at the centre of the sphere so that the upper hemisphere is expressed by the equation

$$z = \sqrt{a^2 - x^2 - y^2};$$

let us take the equatorial circle $x^2 + y^2 = a^2$ as the region D . We readily obtain :

$$\frac{\partial z}{\partial x} = -\frac{x}{\sqrt{a^2 - x^2 - y^2}}, \quad \frac{\partial z}{\partial y} = -\frac{y}{\sqrt{a^2 - x^2 - y^2}},$$

therefore

$$1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 = \frac{a^2}{a^2 - x^2 - y^2},$$

and formula (4) gives us the area of the hemisphere :

$$\frac{S}{2} = a \iint_{x^2 + y^2 \leq a^2} \frac{dx dy}{\sqrt{a^2 - x^2 - y^2}},$$

or, changing to polar coordinates ($x = r \cos \varphi$, $y = r \sin \varphi$)

$$\frac{S}{2} = a \int_0^{2\pi} d\varphi \int_0^a \frac{r dr}{\sqrt{a^2 - r^2}} = 2\pi a \int_0^a \frac{r dr}{\sqrt{a^2 - r^2}} = 2\pi a (\sqrt{a^2 - r^2}) \Big|_a^0 = 2\pi a^2,$$

$$S = 4\pi a^2$$

this formula is well-known from elementary geometry.

For further exercises cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 107, 109, 110, 118.

The established concept of area of a surface has the disadvantage that it depends on the chosen system of coordinates (*i.e.* on the choice of plane onto which the given part of the surface is projected). We could show directly that this dependence is only imaginary (*i.e.* the area remains unchanged even if we alter the direction of projection provided the part of surface is only intersected at one point by every projected straight line); alternatively, we could replace our definition, by another, a more complicated definition which would be completely free from all arbitrary elements. However, both these methods are too complicated to be considered here.

If the part of the surface whose area we are trying to find does not permit the expression of one of its coordinate axes as a single-valued function of the two other axes (such is, for example, every closed surface), then it is frequently possible to divide this part into a finite number of simpler parts so that such an expression is possible (in general, it is necessary in such cases to choose different directions for the projection of individual parts). The area of the whole part is then equal to the sum of the areas of the constituent "simpler" parts.

2. Integrals over parts of surfaces. In § 52 we have defined the length of an arc of a curve and introduced the concept of an "integral along a given section of a curve". Similarly, having defined the area of a part of a curved surface, we can use the concept of "double integral over the given part of the surface" which is naturally a generalisation of the usual double integral over one or other plane region.

Let S be a part of a surface of the type considered at the end of example 1 and let $F(x, y, z)$ a continuous function in a region in space which contains the part S within itself. Let us divide the part

S arbitrarily into cells, each of which has a definite area and choose in every cell σ_k an arbitrary point (x_k, y_k, z_k) . If the sum

$$\sum_k F(x_k, y_k, z_k) \sigma_k$$

tends to a definite limit when the division becomes infinitely fine (*i.e.* when the greatest of the diameters of the cells tends to zero) and if this limit is independent of the performed divisions of the part S and choice of the points (x_k, y_k, z_k) , then we say that this limit is *double integral of the function $F(x, y, z)$ over the given part S of the surface* and denote it by

$$\iint_S F(x, y, z) d\sigma.$$

This concept has many applications which are similar to those of an integral along a given section of a curve. In § 54 we have used this integral to express the mass of a material curve whose density of which at every point is known. Problems involving mass, electric charges, etc. on material parts of surfaces are solved in the same manner. Let us consider, for example, a charged electric conductor whose charge is distributed over its surface with a (surface) density $\rho(x, y, z)$. The reader will have no difficulties in showing for himself that the part S of this surface will have a charge equal to

$$\iint_S \rho(x, y, z) d\sigma.$$

3. Mass of a heterogeneous body. As the simplest example of the use of a triple integral let us determine mass of a heterogeneous physical body with reference to its density. If the given body is homogeneous, *i.e.* its density ρ is the same at every point, then its mass M is equal to the product of the density ρ and the volume of the body V . If the body is heterogeneous, then its density $\rho = \rho(x, y, z)$ is different at different points. Let us divide the given body arbitrarily into cells and let $d(T)$ be the greatest diameter of a cell of the given division T . Let us take an arbitrary cell Δ_k and select an arbitrary point (ξ_k, η_k, ζ_k) in it. We shall assume that the function $\rho(x, y, z)$ is continuous within the given body. If the diameter of the cell Δ_k is very small, the values of the density ρ at different points will be very close to one another and also close to $\rho(\xi_k, \eta_k, \zeta_k)$. It is therefore natural to assume that the mass of the cell Δ_k will be close to the

mass which it would have if it were homogeneous and its density equal to $\rho(\xi_k, \eta_k, \zeta_k)$, *i.e.* it will be close to

$$\rho(\xi_k, \eta_k, \zeta_k) \Delta_k;$$

the mass of the whole body should be close to the sum

$$\sum_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k,$$

taken over all the cells of the body. But we know from § 119 that this sum tends to a definite limit as $d(T) \rightarrow 0$ which we can denote by

$$\iiint_V \rho(x, y, z) dx dy dz,$$

where V is a region in the three-dimensional space occupied by the given body. We naturally take this limit as the expression of the mass M of the given body :

$$M = \iiint_V \rho(x, y, z) dx dy dz.$$

4. Coordinates of centre of gravity and moments of inertia of a body. The apparatus of double integration enables us to find easily the coordinates of centre of gravity and moments of inertia of plane plates while the apparatus of triple integration enables us to do the same for bodies in space. We shall only consider bodies in space, since the arguments and results for plane plates are exactly the same as those which we shall obtain in the three dimensional case.

We shall again divide our body into cells with small diameters and choose an arbitrary point (ξ_k, η_k, ζ_k) in each cell Δ_k . If we replace each cell Δ_k by a material point with the mass

$$\rho(\xi_k, \eta_k, \zeta_k) \Delta_k$$

situated at the point (ξ_k, η_k, ζ_k) , then the given body will be replaced by a system of a finite number of material points whose statical properties will be close to the given body. For such a system of material points the coordinates of centre of gravity will be as follows :

$$\frac{\sum_k \xi_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k}{\sum_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k}, \quad \frac{\sum_k \eta_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k}{\sum_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k}, \quad \frac{\sum_k \zeta_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k}{\sum_k \rho(\xi_k, \eta_k, \zeta_k) \Delta_k}.$$

Therefore we naturally take the limits $\bar{x}, \bar{y}, \bar{z}$ of these three expressions for $d(T) \rightarrow 0$ as the coordinates of centre of gravity which are respectively equal to

$$\bar{x} = \frac{\iiint_V x \rho(x, y, z) dx dy dz}{\iiint_V \rho(x, y, z) dx dy dz}, \quad \bar{y} = \frac{\iiint_V y \rho(x, y, z) dx dy dz}{\iiint_V \rho(x, y, z) dx dy dz},$$

$$\bar{z} = \frac{\iiint_V z \rho(x, y, z) dx dy dz}{\iiint_V \rho(x, y, z) dx dy dz},$$

or denoting by M the mass of the given body

$$\bar{x} = \frac{1}{M} \iiint_V x \rho dx dy dz, \quad \bar{y} = \frac{1}{M} \iiint_V y \rho dx dy dz,$$

$$\bar{z} = \frac{1}{M} \iiint_V z \rho dx dy dz$$

(where for the sake of brevity we write ρ instead of $\rho(x, y, z)$ under the signs of the integrals). In particular, if the given body is homogeneous (*i.e.* ρ is constant throughout the body), then $M = \rho V$ and

$$\bar{x} = \frac{1}{V} \iiint_V x dx dy dz, \quad \bar{y} = \frac{1}{V} \iiint_V y dx dy dz,$$

$$\bar{z} = \frac{1}{V} \iiint_V z dx dy dz.$$

Let us now consider moments of inertia of the given body and again begin with the replacement by the approximate system of a finite number of material points as described above. For such a system moment of inertia with respect to the XOY -plane will be:

$$\sum_k \zeta_k^2 \rho(\xi_k, \eta_k, \zeta_k) \Delta_k$$

and similarly for the two other coordinate surfaces (planes). The moment of inertia with respect to the OX -axis will be:

$$\sum_k (\eta_k^2 + \zeta_k^2) \rho(\xi_k, \eta_k, \zeta_k) \Delta_k$$

and similarly for the other two coordinate planes. Finally moment of inertia of the approximate system with respect to the origin O of coordinates will be:

$$\sum_k (\xi_k^2 + \eta_k^2 + \zeta_k^2) \rho(\xi_k, \eta_k, \zeta_k) \Delta_k.$$

Using the same arguments as those used in finding the coordinates of centre of gravity we find that moment of inertia of the given body with respect to the XOY -plane is equal to:

$$M_{xy} = \iiint_V z^2 \rho(x, y, z) dx dy dz$$

and similarly for the other two coordinate planes. Similarly moment of inertia of the given body with respect to the OX -axis is equal to

$$M_x = \iiint_V (y^2 + z^2) \rho(x, y, z) dx dy dz.$$

Finally moment of inertia of the given body with respect to the origin O of coordinates is equal to:

$$M_o = \iiint_V (x^2 + y^2 + z^2) \rho(x, y, z) dx dy dz.$$

For exercises cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 191, 193, 200, 201.

CHAPTER XXVIII

CURVILINEAR INTEGRALS

§ 121. Definition of a plane curvilinear integral

In chapter 26 we have studied integrals of the type

$$\int_a^b f(x, y) dx,$$

where the variable y is a parameter, *i.e.* it remains constant during integration. Direct generalisation of the part played by the variable y will be the case when, within the limits of integration (*i.e.* when $a \leq x \leq b$), y is given as a function of x , $y = \varphi(x)$, so that the integrand becomes $f[x, \varphi(x)]$. If, as we shall assume, the functions $f(x, y)$ and $\varphi(x)$ are continuous in their corresponding regions, then the function $f[x, \varphi(x)]$ will also be continuous in the interval $a \leq x \leq b$ and there is no doubt as to the existence of the integral

$$\int_a^b f[x, \varphi(x)] dx. \tag{1}$$

Let us denote the beginning and end points of the curve $y = \varphi(x)$ in the interval (a, b) respectively by A and B (fig. 80). In this case the integral (1) is called a *curvilinear integral of the function $f(x, y)$ with respect to x along the curved interval AB* and denoted as follows:

$$\int_{AB} f(x, y) dx; \tag{2}$$

this notation implies that y should be replaced by the function of x whose graph is given by the curvilinear interval AB . Thus if y

retains the constant value y_0 during integration (the case of an integral depending on a parameter), then the integral (2) is taken along the rectilinear interval $y = y_0$ ($a \leq x \leq b$) and the coordinates of the points A and B are respectively equal to (a, y_0) and (b, y_0) .

This concept of curvilinear integral along an interval of a plane curve evidently contains nothing new apart from the fact that we agree to denote the integral (1) by the symbol (2). We must draw attention to the fact that in this notation the direction (from A and B) on the curve AB is important; in fact it follows from the definition of the symbol (2) that

$$\begin{aligned} \int_{BA} f(x, y) dx &= \int_b^a f[x, \varphi(x)] dx = - \int_a^b f[x, \varphi(x)] dx = \\ &= - \int_{AB} f(x, y) dx, \end{aligned}$$

i.e. if the direction of the curve along which we integrate changes, the sign of the integral is reversed.

This initial definition of a curvilinear integral has very limited applications, since very restrictive conditions are placed on the interval AB , *i.e.* y must remain a single-valued function of x along the course of the curve (or, speaking geometrically, any straight line parallel to the OY -axis must intersect this interval at one point only); in practice one must often integrate along interval of much more complicated form; thus in many problems of mechanics and physics it is important to integrate along simple *closed* curves which, of course, do not satisfy the above conditions of the simplest case. We must therefore try to evolve an analytical instrument which would enable us to extend the concept of a curvilinear integral to a wider class of cases.

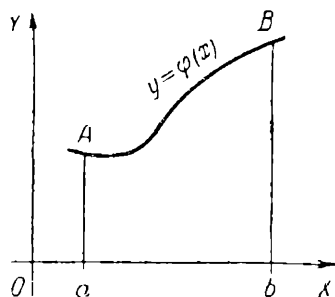


Fig. 80.

Let us then return to the simplest case considered above and assume that the function $\varphi(x)$ represented by the curvilinear interval AB is not only continuous in the interval (a, b) but also has a continuous derivative in that interval. Let us divide the arc AB into subintervals ("cells") by the points of division A_1, A_2, \dots, A_n and denote

the length of the subinterval $A_{k-1} A_k$ of this curve by λ_k . We know (§ 52) that λ_k can be expressed by the integral

$$\lambda_k = \int_{x_{k-1}}^{x_k} \sqrt{1 + \varphi'^2(x)} dx,$$

where x_{k-1} and x_k are respectively the abscissae of the points A_{k-1} and A_k on the curve AB . According to the mean-value theorem we can therefore write

$$\lambda_k = \sqrt{1 + \varphi'^2(\xi_k)} \Delta_k, \quad (3)$$

where $\Delta_k = x_k - x_{k-1}$ and ξ_k is a point in the subinterval (x_{k-1}, x_k) . We also note that since $\varphi'(\xi_k)$ is the tangent of the angle α_k made by the tangent to the curve AB at the point $x = \xi_k$ and the positive direction of the OX -axis, therefore

$$1 + \varphi'^2(\xi_k) = \sec^2 \alpha_k = \frac{1}{\cos^2 \alpha_k}$$

and the relation (3) gives:

$$\lambda_k \cos \alpha_k = \Delta_k,$$

where we must take $\cos \alpha_k > 0$ which implies that α_k *i.e.* the angle between the tangent and the OX -axis, is acute, (or, which is the same, we must direct the tangent towards increasing values of x). This choice of direction of the tangent is necessary if we want $\lambda_k \cos \alpha_k$ to have the same sign as Δ_k since, in accordance with our assumption that $a < b$, we have $\Delta_k = x_k - x_{k-1} > 0$. If we had $a > b$, then we would have $\Delta_k < 0$ for every k ; hence the preservation of the relation $\lambda_k \cos \alpha_k = \Delta_k$ demands that $\cos \alpha_k < 0$; therefore in this case the angle α_k must be acute, *i.e.* we must direct the tangent towards decreasing values of x . Hence in all cases *the direction of the tangent must be chosen in relation to the movement along the curve from A to B*, regardless of whether this movement takes place from left to right or from right to left.

Let us now construct a sum extending over all cells

$$\sum_k f(\xi_k, \eta_k) \cos \alpha_k \cdot \lambda_k = \sum_k f(\xi_k, \eta_k) \Delta_k, \quad (4)$$

where in each term $\eta_k = \varphi(\xi_k)$. We shall make the division of the interval AB sufficiently fine so that the greatest length λ_k (and hence

the greatest length Δ_k) should tend to zero. Since, in accordance with our assumptions, the function $f[x, \varphi(x)]$ is continuous in the interval (a, b) , therefore the right-hand side of the equation (4) will tend to the following integral as to its limit:

$$\int_a^b f[x, \varphi(x)] dx,$$

which we agree to denote by the symbol

$$\int_{AB} f(x, y) dx$$

and call as curvilinear integral of the function $f(x, y)$ along the curve AB . Hence the left-hand side of the equation (4) also tends to this limit, *i.e.* we have:

$$\int_{AB} f(x, y) dx = \lim \sum_k f[\xi_k, \varphi(\xi_k)] \cos \alpha_k \cdot \lambda_k. \quad (5)$$

If, in general, we agree to denote by $\alpha = \alpha(x, y)$ the angle between the tangent at the point (x, y) to the curve AB and the positive direction of the OX -axis, we shall have $\alpha_k = \alpha(\xi_k, \eta_k)$; the sum on the right-hand side of the equation (5) can then be written in the form

$$\sum_k f(\xi_k, \eta_k) \cos \alpha(\xi_k, \eta_k) \cdot \lambda_k = \sum_k F(\xi_k, \eta_k) \lambda_k,$$

where $F[x, \varphi(x)] = f[x, \varphi(x)] \cos \alpha[x, \varphi(x)]$ is a continuous function of x in the interval (a, b) . But, as a result of our assumption on continuity of the derivative $\varphi'(x)$, the curve $y = \varphi(x)$ can be extended along the interval (a, b) and for such curves the sum

$$\sum_k F(\xi_k, \eta_k) \lambda_k,$$

where (ξ_k, η_k) is an arbitrary point on the line λ_k , tends to a definite limit when the division becomes infinitely fine (§ 52); we agree to denote this limit by

$$\int_{AB} F(x, y) d\lambda$$

and call it integral of the function $F(x, y)$ along the curve AB .

Thus in our case

$$\sum_k f[\xi_k, \varphi(\xi_k)] \cos \alpha[\xi_k, \varphi(\xi_k)] \lambda_k \rightarrow \int_{AB} f(x, y) \cos \alpha(x, y) d\lambda.$$

Comparing this with the relation (5) we obtain :

$$\int_{AB} f(x, y) dx = \int_{AB} f(x, y) \cos \alpha(x, y) d\lambda. \quad (6)$$

The left-hand side of this equation is a curvilinear integral along the interval AB in the sense defined at the beginning of this paragraph. The right-hand side represents an "integral along the curve AB " as defined in § 52. The latter definition differs essentially from the former, for it is constructive and the integral defined by it results from a definite construction *i.e.* it is a sum of a definite form. We can therefore regard the relation (6) as a new constructive definition of a curvilinear integral. Let us note further that $\alpha(x, y)$ denotes here the angle between the positive direction of the OX -axis and the tangent to the curve AB at the point (x, y) drawn in the direction of movement from A to B (*i.e.* $\cos \alpha(x, y) > 0$ for $a < b$ and $\cos \alpha(x, y) < 0$ for $a > b$).

So far all that was said only referred to the simplest case when the given curve was expressed along the interval AB by the equation

$y = \varphi(x)$. If this is not so, the integral on the left-hand side of the equation (6) is meaningless, for then each value of x on the curve AB corresponds, in general, to several values of y . However, the position is quite different for the integral on the right-hand side of this equation. The constructive definition given to this integral in § 52 is independent of the fact that the curve AB can be represented by an equation of the form $y = \varphi(x)$; it

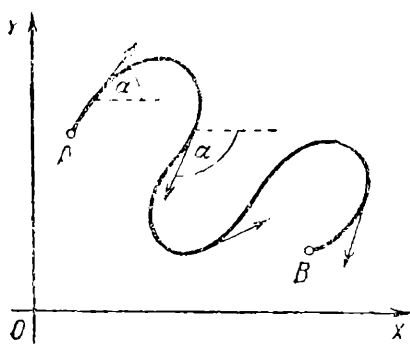


Fig. 81.

remains valid in more general cases and, in particular, it holds for all "smooth" curves AB . In these general cases the direction from A to B in which the curve is described will, in general, no longer be always from left to right (or always from right to left); in different intervals of the curve AB this direction can vary (fig. 81) so that the

angle $\alpha(x, y)$ can be acute and obtuse and $\cos \alpha(x, y)$ will be correspondingly positive and negative. For the integrand on the right-hand side of the equation (6) we draw, as before, a tangent from A to B in order to define the angle α at every point.

These considerations naturally lead us to *define* a curvilinear integral

$$\int_{AB} f(x, y) dx$$

by the equation (6) in all cases in case the integral on the right-hand side of this equation exists; we have just shown that this extension of the concept of curvilinear integral enables us to integrate along a wider class of curves which also includes some of the simple closed curves, e. g. circles, ellipses, etc.

Taking the general definition of a curvilinear integral as given by formula (6) we identify the concept of a curvilinear integral with the concept of an "integral along an extended curve AB " as given in § 52. However, this is not quite so. In fact, in our former definition of the integral

$$\int_C F(x, y) d\lambda,$$

taken along the extended curve C , the integrand only depended on the point (x, y) on the curve C ; in our case the integrand $f(x, y) \cos \alpha(x, y)$ on the right-hand side of formula (6), apart from depending on x and y , also depends essentially on the direction of the tangent to the curve AB at the point (x, y) . If this direction is changed, $\cos \alpha(x, y)$, and hence the integrand as a whole, reverses its sign. The former definition of an "integral along the curve C " is independent of the direction in which the curve C is described; in fact, according to this definition, the choice of this direction is quite irrelevant for this integral. For the integral on the right-hand side of the equation (6) the position is quite different. Having established a definite direction on the curve AB , say from A to B , we simultaneously determine a definite direction at every point on the tangent to this curve; thus $\cos \alpha(x, y)$, and hence the integrand as a whole, receives a definite value at the point (x, y) and our integral becomes an "integral along the curve AB ". If we change the direction along the curve

AB , then the integrand changes its sign at every point and therefore the sign of the integral will also be reversed :

$$\int_{BA} f(x, y) dx = - \int_{AB} f(x, y) dx,$$

and the reason for this difference of signs is, as we can see, due to the fact that formula (6) defines $\int_{AB} f dx$ and $\int_{BA} f dx$ by means of two differ-

ent "integrals along the curve AB ".

In curvilinear integrals the curve along which we integrate is frequently denoted by a single letter, for example by C ; but this notation does not show us the chosen direction on this curve and in such cases this direction must always be discussed, for otherwise the integral has no definite meaning. Thus if the curve C has ends A and B , it is usually denoted as AB or BA in accordance with the chosen direction of integration. If the curve C is closed and has no ends then in simpler cases it divides the plane into two parts — the interior and exterior. In such cases we call the *direct way* when direction of describing the curve where the interior always remains to the left of the path and the *reverse way* when the curve is described in the opposite direction. If the curvilinear integral is denoted by

$$\int_C F(x, y) dx,$$

where C is a closed curve, then in cases where nothing is said about direction it is assumed that a direct description of the curve is envisaged ; we shall keep to this rule in future.

It is self-evident that all that is said in this paragraph in relation to integrals of the form

$$\int_{AB} f(x, y) dx,$$

can be extended without modifications to integrals of the form

$$\int_{AB} f(x, y) dy.$$

If the interval AB along which we integrate can be expressed by an equation of the type

$$x = \psi(y) \quad (c \leq y \leq d),$$

we assume by definition

$$\int_{AB} f(x, y) dy = \int_c^d f[\psi(y), y] dy.$$

Similarly to formula (6) we then prove that in this "simple" case

$$\int_{AB} f(x, y) dy = \int_{AB} f(x, y) \sin \alpha d\lambda \quad (7)$$

(since for the *acute* angle β between the tangent to the curve AB and the OY -axis we evidently have $\cos \beta = \sin \alpha$) and we take this formula as the definition of its left-hand side in every case when its right-hand side has a meaning. It is evident that in this case the direction of the tangent at every point (and hence also the sign of $\sin \alpha$) is chosen in accordance with the direction of the curve AB itself.

In practice one frequently meets sums of the form

$$\int_{AB} P(x, y) dx + \int_{AB} Q(x, y) dy,$$

whose terms differ from one another by their integrands and variables of integration but where the integrals are taken along the same curve AB . In such cases it is customary to write the sum in the form

$$\int_{AB} [P(x, y) dx + Q(x, y) dy],$$

or, more briefly,

$$\int_C (P dx + Q dy).$$

It follows from the formulae (6) and (7) that

$$\int_C (P dx + Q dy) = \int_C (P \cos \alpha + Q \sin \alpha) d\lambda. \quad (8)$$

For exercises to § 121 *cf.* Problem Book by B.P. Demidovich, Section VIII, Nos. 293, 295, 299.

§ 122. Work of a plane field of force

Curvilinear integrals find many applications in geometry, physics and technical problems. Before proceeding further we shall consider one of the most important and at the same time most typical examples of this kind.

In § 45 we have already considered the work done by a variable force when a material point moves along a rectilinear path; we shall now consider the same problem on the assumption that the point moves along a plane curve.

Let a material point situated at the point (x, y) in the given plane be subjected to the action of the force F whose magnitude and direction are uniquely defined by the coordinates x and y of the point at which it is situated. Thus a definite vector F which expresses the force acting on the material point situated at the given spot is associated to every point in the plane (or each part of the plane). The set of these vectors is known as the *plane field of force* defined in the given plane (or in a definite part of it).

Let us now assume that the material point travels along the interval AB of a given smooth curve in the plane in which the field of force F is given. We are trying to find the work of this displacement done by forces acting in this field. In order to define a field of force it is very convenient to define the components $P(x, y)$ and $Q(x, y)$ of the vector (force) F in the direction of the axes OX and OY ; we assume that these two functions are continuous in the region which contains the curve AB within itself.

Let us divide the route AB of the material point into sub-intervals (cells) $\lambda_1, \lambda_2, \dots, \lambda_n$ and choose an arbitrary point $A_k(x_k, y_k)$ in every cell λ_k . Let F_k denote the vector (force) acting at the point A_k and $|F_k|$ the absolute value of this vector. Let φ_k denote the angle between the vector F_k and the tangent to the curve AB at the point A_k taken in the direction of motion (or, which is the same, the velocity vector of the moving point at the point A_k). If the sub-interval λ_k has the same length but is rectilinear and follows the direction of the above tangent and the force acting in this subinterval is constant and equal to the vector F_k , then the work done by this force in passing through the given point in the interval λ_k would, in accordance with the law of elementary physics, be equal to the product

$$|F_k| \cos \varphi_k \cdot \lambda_k.$$

If the cell λ_k is small while the vector F and the direction of the tangent to the curve AB change continuously as the positions of the given point change, then the above product gives an approximate expression of the work of the field done in displacing a moving point along the subinterval λ_k . And since we naturally assume that the work of the field along the whole length of the curve AB is equal to the sum of the ‘elementary’ works along its individual subintervals, we obtain the following expression which gives the approximate value of this sum

$$\sum_{k=1}^n |F_k| \cos \varphi_k \cdot \lambda_k,$$

this approximation will be more accurate if the dimensions (diameters) of the cells λ_k become smaller. We know (§ 52) that if the division of the interval AB becomes indefinitely fine, this sum tends to a definite limit which is independent of the divisions and choice of the points (x_k, y_k) in the subintervals; we naturally take this limit as the exact expression of the work of the field along the interval AB of the given curve. Under these conditions we shall agree to denote the limit of this sum by

$$\int_{AB} |F| \cos \varphi d\lambda,$$

where $|F| \cos \varphi$ denotes the projection of the vector (force) acting at the point (x, y) on the curve AB on to the direction of the path at that point. But in accordance with vector algebra the projection of the vector in any direction is equal to the sum of the projections of its components in the same direction; therefore if we denote by $\alpha = \alpha(x, y)$ the angle between the positive direction of the OX -axis and the velocity vector of the moving point at the point (x, y) we obtain:

$$|F| \cos \varphi = P(x, y) \cos \alpha + Q(x, y) \sin \alpha,$$

which gives us the following expression for the work of the field along the interval AB :

$$W = \int_{AB} \{P(x, y) \cos \alpha + Q(x, y) \sin \alpha\} d\lambda,$$

or, in accordance with formula (8) § 121

$$W = \int_{AB} \{P(x, y) dx + Q(x, y) dy\},$$

which fully solves our problem. We thus see that the physical concept connected with this problem can be accurately formulated in mathematics by means of the general concept of a curvilinear integral.

§ 123. Green's formula

In practice curvilinear integrals round closed curves are very important; in this and the next paragraphs we shall pay particular attention to such integrals. The regions of integration will in all cases be smooth curves, *i.e.* curves which do not intersect themselves; such curves, as we already know, always divide a plane into two parts—the exterior and the interior. Denoting such a contour by L we shall take the integral

$$\int_L (P dx + Q dy),$$

as is usually accepted, to denote an integral in which the curve L is described the “direct” way, *i.e.*, such that the interior of the plane always remains to the left of the direction of movement (fig. §2). The opposite direction is called the “reverse” way.

Let us consider a region D in the XOY -plane bounded from above and below by the curves $y = \varphi_1(x)$ and $y = \varphi_2(x)$ respectively and from sides by the straight lines $x = a$ and $x = b$ (fig. 83). Both straight lines (or one of these lines) can degenerate into points so that the curves $y = \varphi_1(x)$ and $y = \varphi_2(x)$ meet at $x = a$ and $x = b$; we shall assume that the functions $\varphi_1(x)$ and $\varphi_2(x)$ are continuous in the interval (a, b) . Let $P(x, y)$ be a continuous function with continuous partial derivatives inside and on the boundary of the region D . We know from the previous chapter that under these circumstances the double integral

$$\iint_D \frac{\partial P}{\partial y} dx dy$$

exists and can be represented in the form

$$\int_a^b \left\{ \int_{\varphi_2(x)}^{\varphi_1(x)} \frac{\partial P}{\partial y} dy \right\} dx.$$

The inner integration (with respect to y) is, as usual, carried out on the assumption that x remains constant; in this case we evidently have

$$\int_{\varphi_2(x)}^{\varphi_1(x)} \frac{\partial P(x, y)}{\partial y} dy = P[x, \varphi_1(x)] - P[x, \varphi_2(x)],$$

and we obtain:

$$\iint_D \frac{\partial P}{\partial y} dx dy = \int_a^b P[x, \varphi_1(x)] dx - \int_a^b P[x, \varphi_2(x)] dx.$$

In accordance with the initial definition in § 121, each integral on the right-hand side represents a curvilinear integral of the function

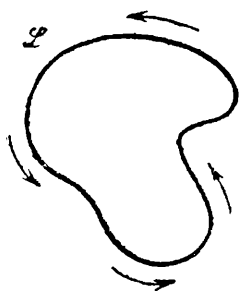


Fig. 82

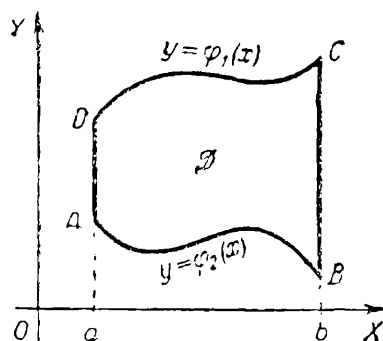


Fig. 83

$P(x, y)$; the first integral is taken along the interval DC of the curve $y = \varphi_1(x)$ and the second along the interval AB of the curve $y = \varphi_2(x)$. We can therefore write

$$\begin{aligned} \iint_D \frac{\partial P}{\partial y} dx dy &= \int_{DC} P(x, y) dx - \int_{AB} P(x, y) dx = \\ &= - \int_{CD} P(x, y) dx - \int_{AB} P(x, y) dx. \end{aligned} \quad (1)$$

Let us now note that integration of the function $P(x, y)$ with respect to x along any one of the rectilinear intervals AD and BC (in any direction) gives zero; in fact, the integrals along these intervals cannot be defined in the sense given in § 121, since y is not a single-valued function of x in these intervals; but the wider definition of § 121 can be applied and it gives zero for both integrals since evidently $\cos \alpha = 0$ in the whole length of each interval.

With this in consideration we can rewrite formula (1) in the form

$$\iint_D \frac{\partial P}{\partial y} dx dy = - \int_{AB} P dx - \int_{BC} P dx - \int_{CD} P dx - \int_{DA} P dx,$$

or

$$\iint_D \frac{\partial P}{\partial y} dx dy = - \int_L P(x, y) dx, \quad (2)$$

where L is the contour of the region D which, in accordance with our agreement, is described in the direct way.

Let us now imagine that the variables x and y are interchanged in all the above arguments; in that case instead of the shape represented in fig. 83 the region D will have the shape shown in fig. 84. Let $Q(x, y)$ be a continuous function with continuous partial derivatives in this new region D . The reader should perform all calculations in full analogy to the above and prove the following relation:

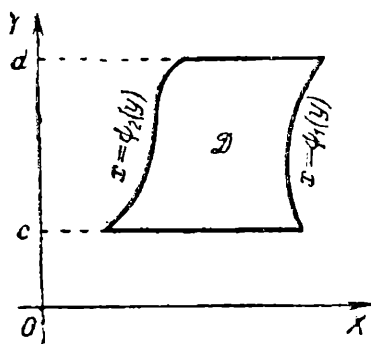


Fig. 84

$$\iint_D \frac{\partial Q}{\partial x} dx dy = \int_L Q(x, y) dy, \quad (3)$$

which is analogous to the relation (2) but differs from it in the sign of the right-hand side. This difference which may at first appear strange is due to the fact that in the choice of a definite direction to describe closed curves the mutual position of the coordinate axes loses its symmetry; by describing a circle with centre at origin of coordinates in the direct way (fig. 85), we must in transit from the positive direction of the OX -axis to the positive direction of the OY -axis describe an arc of 90° whereas the reverse description involves an arc of 270° (or an arc of 90° in the opposite direction).

Let us now assume that the shape of the region D is, as shown in fig. 83 and fig. 84 (this condition is satisfied by every circle, ellipse, rectangle and more general figures of the type shown in fig. 86). If $P(x, y)$, $Q(x, y)$ are continuous functions with continuous partial derivatives inside and on the boundary of the region D , then both relations (2) and (3) hold.

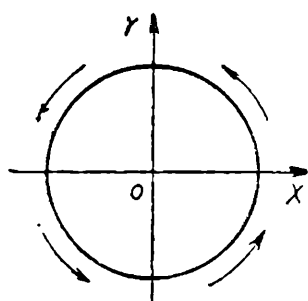


Fig. 85

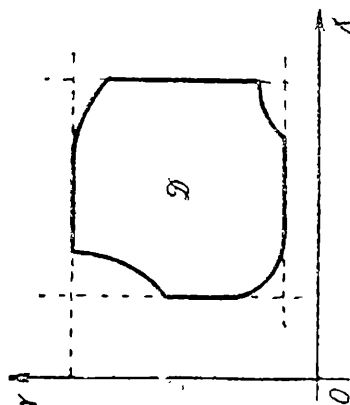


Fig. 86

Subtracting (2) from (3) we obtain :

$$\int_D \int \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_L (P dx + Q dy). \quad (4)$$

This most important relation which connects a double integral with a curvilinear integral and has numerous applications is usually known as *Green's formula*. We have proved

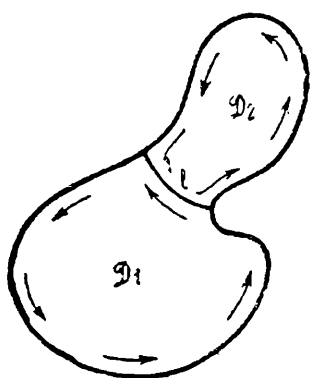


Fig. 87

this formula for regions of specific form. However, Green's formula can be readily extended to a much wider class of regions. In fact, let us consider a region D bounded by a simple (smooth, *i.e.* not intersecting itself) contour (Fig. 87) and draw a smooth curve l to divide this region into two regions D_1 and D_2 . Assuming that Green's formula holds in both regions D_1 and D_2 , we can write it for these regions and add the two formulae term-by-term. We then obtain on the left-hand

side the double integral

$$\int_D \int \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy,$$

while on the right-hand side we have a sum of two curvilinear integrals taken along the contours of the regions D_1 and D_2 respectively in the direct way. The arrows in fig. 87 show that these integrals describe the curve l in opposite directions as a result of which their corresponding parts cancel each other; the remaining parts of the integrals evidently give one integral taken along the contour L of the region D in the direct way; we thus see that if Green's formula holds in the regions D_1 and D_2 , it also holds in the region D . By repeating this argument we readily find that this

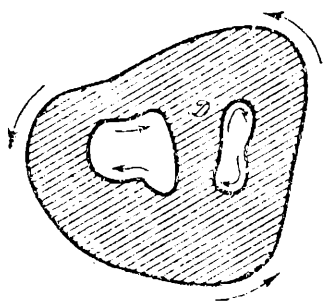


Fig. 88

result remains valid if the drawn lines divide the region D into an arbitrary number of regions, provided Green's formula holds in each region. Hence this formula holds not only in regions of specific form for which we have proved it above but also in every other region which can be divided into a finite number of regions of this kind by means of corresponding curves. And as can readily be seen, this gives a very wide class of plane regions. Green's formula also holds in the so-called "multiply-connected" regions with "holes" (fig. 88), *i.e.* in regions which are bounded not by a single curve but by several closed curves, provided this region is the sum of regions of the simple type considered above. The curvilinear integral on the right-hand side of Green's formula then represents the sum of integrals round the outer contour of the region D and the contours of all the "holes", where each contour is described in the direct way, *i.e.* such that the region D always remains to the left of the direction in which the contour is described.

Finally it can be shown that Green's formula also holds in every region bounded by a smooth closed curve. For this purpose we must begin by proving that the region bounded by a polygon (broken contour) is always equal to the sum of regions of the type considered above so that Green's formula can be applied to every polygon. After this we inscribe in the given polygon another polygon with very small sides and apply Green's formula to it; assuming that the sides of this polygon become infinitely small we can show by a limiting process that the relation expressed by Green's formula also remains valid for the given region bounded by a smooth contour. We cannot give details of this proof here.

For exercises to § 123, cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 341, 342.

§ 124. Application to differentials of functions of two variables

Apart from being theoretically important Green's formula finds many applications in problems of physics and mathematical analysis. One of the most important applications of this kind is considered in this paragraph.

Let $P(x, y)$ and $Q(x, y)$ again denote two continuous functions with continuous partial derivatives in a region D in the xy -plane. For the sake of simplicity we shall assume in all that follows that the region D and all other regions which we shall meet in this paragraph are open (i.e. consist of interior points only) that they are connected (i.e. they have no "holes") and bounded by simple (non-intersecting) smooth contours.

We know (§ 90) that if the expression

$$P(x, y) dx + Q(x, y) dy \quad (1)$$

is differential of a function $F(x, y)$ in a region D , then in this region

$$P(x, y) = \frac{\partial F}{\partial x}, \quad Q(x, y) = \frac{\partial F}{\partial y}; \quad (2)$$

conversely, as a result of the assumed continuity of the functions P and Q , it follows from the relations (2) that the function $F(x, y)$ has a differential in the region D which is expressed by formula (1). It is obvious that for some analytical problems it is essential to have a criterion in order to determine whether the expression (1) will be differential of a function of two variables in the given region. We shall see that Green's formula enables us to establish the necessary and sufficient condition for this purpose which must be satisfied by the functions P and Q ; in fact, we shall deduce not one but three such necessary and sufficient conditions (which are, of course, equivalent but expressed in different terms).

Theorem. For each of the following four statements:

1° the expression $P dx + Q dy$ is a differential of a function $F(x, y)$ in the region D ;

2° we have $\partial P / \partial y = \partial Q / \partial x$ everywhere in region D ;

3° the curvilinear integral

$$\int (P dx + Q dy),$$

round any smooth closed curve which lies entirely within the region D is equal to zero;

4° the curvilinear integral

$$\int_{AB} (P dx + Q dy)$$

only depends on the points A and B in the region D and not on the path joining these points along which it is taken, provided this path is a smooth curve and lies entirely within the region D ;

— the other three are corollaries.

This theorem also shows that any of the three conditions 2°, 3° and 4° can be the necessary and sufficient condition for the expression $P dx + Q dy$ to be a differential of the function $F(x, y)$ in the region D ; we shall also see that the proof of this theorem will enable us to find an expression for the function $F(x, y)$ in all cases when it exists.

In order to prove this theorem we shall begin by establishing four auxiliary propositions whose set will be equivalent to this theorem.

Lemma 1. 2° follows from 1°.

Proof. It follows from $P dx + Q dy = dF$ that

$$P = \frac{\partial F}{\partial x}, \quad Q = \frac{\partial F}{\partial y};$$

therefore

$$\frac{\partial P}{\partial y} = \frac{\partial^2 F}{\partial x \partial y}, \quad \frac{\partial Q}{\partial x} = \frac{\partial^2 F}{\partial y \partial x};$$

since the functions $\partial P / \partial y$ and $\partial Q / \partial x$ are continuous, therefore the righthand sides of these equations are equal to one another; hence the left-hand sides are also equal and lemma 1 is proved.

Lemma 2. 3° follows from 2°.

Proof. Let L be a closed smooth curve which lies entirely within the region D . Let us apply Green's formula to the region

Δ bounded by the curve L . Since, according to our assumption, we have $\partial P / \partial y = \partial Q / \partial x$ everywhere in the region D , therefore the left-hand side of Green's formula vanishes and we have:

$$\int_L (P dx + Q dy) = 0,$$

which was to be proved.

Lemma 3. 4° follows from 3° .

Proof. Let us join the points A and B in the region D by two arbitrary smooth curves L_1 and L_2 which lie entirely within the region D and do not intersect one another (fig. 89). In that case the curve L_1 from A to B and the curve L_2 from B to A together form a closed curve; hence as a result of our assumptions we have

$$\int_{AB}^{(L_1)} (P dx + Q dy) + \int_{BA}^{(L_2)} (P dx + Q dy) = 0,$$

and since

$$\int_{BA}^{(L_2)} (P dx + Q dy) = - \int_{AB}^{(L_2)} (P dx + Q dy),$$

therefore

$$\int_{AB}^{(L_1)} (P dx + Q dy) = \int_{AB}^{(L_2)} (P dx + Q dy);$$

this proves lemma 3 for non-intersecting curves. However, if the curves L_1 and L_2 intersect one another (fig. 90), we must join the

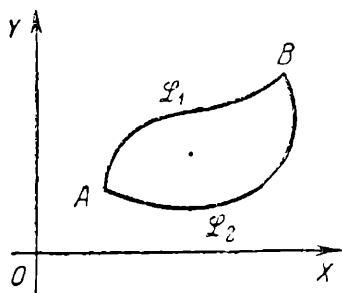


Fig. 89

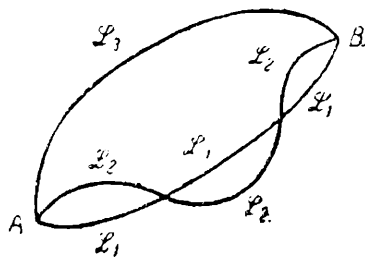


Fig. 90

points A and B by a third curve L_3 which also lies in the the region D and does not intersect the curve L_1 or the curve L_2 *). It then follows from our proof that the integral along the curve L_3 will coincide with the integral along the curve L_1 and the integral along the curve L_2 ; hence these two integrals must coincide with one another and lemma 3 is proved.

Lemma 4. 1° follows from 4°.

Proof. Let us consider a fixed point $A (x_0, y_0)$ and a variable point $B (x, y)$ in the region D , where the point B can move anywhere in the region D . Since we are assuming that the integral

$$\int_{AB} (P dx + Q dy)$$

is independent of the curve joining the points A and B along which it is taken (provided this curve is smooth and lies entirely within the region D), therefore, as the position of the point A is fixed, this integral only depends on the position of the point B , i.e. it is a single-valued function $F (x, y)$ of its coordinates. We will show that at every point of the region

$$\frac{\partial F}{\partial x} = P, \quad \frac{\partial F}{\partial y} = Q,$$

and hence, as a result of continuity of the functions P and Q

$$dF = P dx + Q dy,$$

which proves lemma 4.

Let us prove, for example, the relation $\partial F / \partial x = P$. Let $B (x, y)$ be an arbitrary (interior) point in the region D ; if $|h|$ is

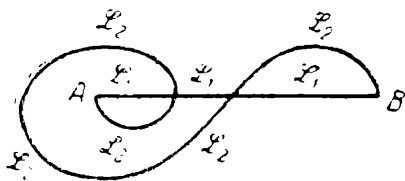


Fig. 91

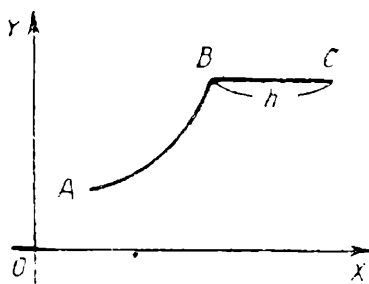


Fig. 92

*) We are assuming here, for the sake of brevity, that this curve L_3 exists. However, this is not always so as can be seen from the case depicted in fig. 91.

sufficiently small, then the point $C(x + h, y)$ also lies in the region D . Let us join the points A and B by an arbitrary smooth curve which lies entirely within the region D ; continuing this curve to the point C with the help of the straight line BC we evidently obtain the (smooth) curve ABC which joins the points A and C and lies entirely within the region D (fig. 92). We therefore have:

$$\begin{aligned} F(x, y) &= \int_{AB} (P dx + Q dy), \\ F(x + h, y) &= \int_{ABC} (P dx + Q dy) = \\ &= \int_{AB} (P dx + Q dy) + \int_{BC} (P dx + Q dy), \end{aligned}$$

hence

$$F(x + h, y) - F(x, y) = \int_{BC} (P dx + Q dy) = \int_{BC} P dx,$$

since *)

$$\int_{BC} Q dy = 0.$$

But it follows from the mean-value theorem that

$$\int_{BC} P dx = h P(\xi, y),$$

where $x < \xi < x + h$. We therefore obtain:

$$\frac{F(x + h, y) - F(x, y)}{h} = P(\xi, y);$$

and since we have $\xi \rightarrow x$ as $h \rightarrow 0$ and the function P is continuous at the point $B(x, y)$, therefore

$$\frac{\partial F}{\partial x} = P(x, y);$$

*) Since the curve BC is horizontal; this can be shown formally in the same way as in the deduction of Green's formula where the curvilinear integral with respect to x along a vertical section was shown to be equal to zero.

the following relation is proved similarly

$$\frac{\partial F}{\partial y} = Q(x, y),$$

which concludes the proof of lemma 4.

The lemmas 1 — 4 are evidently equivalent to the fundamental theorem formulated above, and its proof is thus also concluded.

For exercises to § 124, cf. Problem Book by B. P. Demidovich, Section VIII, Nos. 300, 302, 305, 308.

§ 125. Curvilinear integrals in space

The concept of curvilinear integrals can be extended without difficulties to cases when integration is carried out along curves in space. In the simplest case when the given section AB of a curve in space can be expressed by equations of the form

$$y = \varphi(x), z = \psi(x) \quad (a \leq x \leq b \text{ and } a \geq x \geq b),$$

where the functions $\varphi(x)$ and $\psi(x)$ are continuous in the interval (a, b) , we assume, in accordance with the definition, that

$$\int_{AB} F(x, y, z) dx = \int_a^b F[x, \varphi(x), \psi(x)] dx;$$

this implies that as in the case of plane curves

$$\int_{BA} F(x, y, z) dx = - \int_{AB} F(x, y, z) dx.$$

This initial definition only applies to intervals of very simple form but it can be extended in the same way as in plane curves.

Assuming that the functions $\varphi(x)$ and $\psi(x)$ have continuous derivatives in the interval (a, b) we again divide the arc AB into "cells" λ_k which are now small arcs of a curve in space. We have (§ 53) :

$$\lambda_k = \int_{x_{k-1}}^{x_k} \sqrt{1 + \varphi'^2(x) + \psi'^2(x)} dx,$$

where x_{k-1} and x_k are the abscissae of the ends of the cell λ_k .

Let us denote by $\alpha = \alpha(x, y, z)$ the angle between the tangent to the curve AB at the point (x, y, z) and the positive direction of the OX -axis (where the direction of the tangent is chosen in relation to the movement from A to B); we therefore have (§ 98)

$$\cos \alpha = \pm \frac{1}{\sqrt{1 + \varphi'^2(x) + \psi'^2(x)}},$$

where we have “+” in the interval λ_k provided $x_k > x_{k-1}$ and “−” in the opposite case. By repeating the arguments used in § 121 word by word for plane curves we readily arrive at the conclusion that in the case of a *simple* interval AB the following formula holds :

$$\int_{AB} F(x, y, z) dx = \int_{AB} F(x, y, z) \cos \alpha(x, y, z) d\lambda,$$

which is analogous to formula (6) § 121. Here, as before, the integral on the right-hand side is defined constructively (as the limit of a sum of definite form) and also holds along an interval of more complicated form than that of the interval AB (thus it always holds when these intervals can be divided into a finite number of simple subintervals; this embraces the simple closed curves; the integral evidently also has a sense in the more general case when AB is an arbitrary smooth curve as we have shown in § 121 for plane curves). Hence we can in this case also regard the formula obtained as definition of the space integral on its left-hand side. In simple cases this definition always coincides with the earlier definition.

It is clear that all we have said above refers to integration with respect to the variables y and z . If P , Q and R are three continuous functions of x , y and z in a region in space which contains the curve $L = AB$ within itself, then similar to formula (8) § 121 we have :

$$\int_L (P dx + Q dy + R dz) = \int_L (P \cos \alpha + Q \cos \beta + R \cos \gamma) d\lambda,$$

where α , β and γ are the angles between the tangent to the curve L at the point (x, y, z) and the positive direction of the axes of coordinates; the direction of the tangent must be taken in relation to the chosen direction of movement along the curve L .

Owing to the fact that physical processes connected with the existence of fields of force usually take place not in a plane but in

three dimensional space, therefore, it is clear why it is desirable to calculate the work performed by the forces of the field in displacing a point along any smooth curve in space (assuming, of course, that the field itself acts in space or in a part of it). If the vector of the field at the point (x, y, z) of the three-dimensional space is given in terms of its components $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$, we can show in the same way as in § 122 where we have dealt with plane fields, that the work W done by the field in displacing a point along the curve AB is expressed by the curvilinear integral

$$\int_{AB} (P dx + Q dy + R dz).$$

Finally as in the plane case, curvilinear integrals along curves in space are important in deciding whether the expression

$$P(x, y, z) dx + Q(x, y, z) dy + R(x, y, z) dz$$

is differential of a function $F(x, y, z)$ of three variables. For regions in space, whose form is sufficiently simple, a proposition holds completely analogous to the theorem in § 124, and involving equivalence of the following four propositions :

1°' *The expression $P dx + Q dy + R dz$ is differential of a function $F(x, y, z)$ in the region V ;*

2°' *the following equations hold everywhere in the region V*

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x}, \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y};$$

3°' *the curvilinear integral*

$$\int (P dx + Q dy + R dz)$$

taken along any smooth closed curve which lies entirely within the region V is equal to zero ;

4°' *the curvilinear integral*

$$\int_{AB} (P dx + Q dy + R dz)$$

depends only on the points A and B and not on the curve connecting these points along which it is taken, provided this curve is smooth and lies entirely within the region V .

In order to prove this equivalence we must prove four lemmas which are completely analogous to the lemmas 1—4 § 124. The formulation of these lemmas is quite clear and we shall not give them here. The reader will have no difficulty in showing that propositions analogous to the lemmas 1, 3 and 4, § 124 are proved in exactly the same way as before ; however, the position is somewhat different with the proposition analogous to lemma 2 : at present we cannot prove that $3''$ follows from $2''$ as this is done for lemma 2 § 124, since we have no formula for curvilinear integrals in space analogous to Green's formula for plane integrals. We cannot deduce such a formula in this chapter since a new concept of *surface integrals* is involved which we shall introduce in the next chapter. We shall return to this problem later and on the basis of a formula, similar to Green's formula, which applies to curvilinear integrals in space, we shall conclude the proof of equivalence of the statements $1''$ — $4''$ by proving the remaining link in the chain of our lemmas (" $3''$ follows from $2''$ ") which we are unable to do here.

For exercises to § 125 cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 326, 328, 335.

CHAPTER XXIX

SURFACE INTEGRALS

§ 126. The simplest case

In § 121 we have defined a curvilinear integral by developing the concept of an integral depending on a parameter (or, in the case of integrals in space, on two parameters). If we begin with a double integral depending on a parameter, a similar development will lead us directly to the important concept of surface integrals to which this chapter is devoted.

Let us consider a surface of x and y in a region D (we can at present assume that this region D is any arbitrary measurable figure) which is expressed by the equation

$$z = f(x, y), \quad (1)$$

where the function $f(x, y)$ is continuous in the region D . Let us denote by S a part of the surface (1) projected onto the region D by means of straight lines parallel to the OZ -axis. Let $F(x, y, z)$ be a continuous function in a given region of the three-dimensional space containing the whole part S of the surface (1). If we assume that the number z is constant, the double integral

$$\iint_D F(x, y, z) \, dx \, dy$$

will contain z as a parameter and evidently be a function of this parameter. However, we can adopt a more general point of view and assume that z is an arbitrary function of x and y and is continuous in the region D ; let this function be, say, the function $f(x, y)$; in that case the given integral takes the form

$$\iint_D F[x, y, f(x, y)] \, dx \, dy,$$

where the integrand is continuous in the region D so that there is no doubt as to the existence of the integral. Such an integral is called a *surface integral* of the function $F(x, y, z)$ on the part S of the surface (1) and denoted by

$$\iint_S F(x, y, z) dx dy;$$

the index S on the integral indicates that z must be regarded as a function of the variables x and y during integration and defined by the equation (1) of the surface, of which S is a part. Thus if z is constant in the region D (the initial part in our development), this implies that the surface (1), over a part of which we integrate, is a plane parallel to the XOY -plane and the surface integral becomes the usual double integral.

Let us now assume that the function $f(x, y)$ is continuous in the region D and has continuous partial derivatives in this region. Let us divide the part S of the given surface into smaller parts ("cells") with small diameters*). The area σ_k of the cell with the same index can be expressed by the following integral as shown in § 120:

$$\sigma_k = \int_{\Delta_k} \int \frac{dx dy}{\cos \gamma} = \int_{\Delta_k} \int \sqrt{1 + f'^2_x(x, y) + f'^2_y(x, y)} dx dy,$$

where Δ_k denotes the projection of the cell σ_k onto the XOY -plane and γ is the acute angle between the normal to the given surface at the point $[x, y, f(x, y)]$ and the positive direction of the OZ -axis. In accordance with the mean-value theorem (§ 116) we have:

$$\sigma_k = \frac{\Delta_k}{\cos \gamma_k},$$

where γ_k is the value of the angle γ at the point (x_k, y_k, z_k) of the cell σ_k .

Let us now construct the sum

$$\sum_k F(x_k, y_k, z_k) \cos \gamma_k \cdot \sigma_k = \sum_k F[x_k, y_k, f(x_k, y_k)] \Delta_k$$

*) So as not to complicate our arguments with details which have no direct connection with the subject under consideration, we shall always assume in future that these cells, and all other regions which we shall encounter, are connected figures bounded by relatively simple contours so that we can apply the concepts and results obtained above to all of them.

including all cells on the part S ; since, according to our assumptions, $F[x, y, f(x, y)]$ is a continuous function of x and y in the region D , therefore the right-hand side of this equation, in accordance with the result of chapter 27, will tend to the following double integral as its limit when the above division becomes indefinitely fine:

$$\iint_D F[x, y, f(x, y)] dx dy,$$

which is no other than the surface integral defined above

$$\iint_S F(x, y, z) dx dy.$$

This surface integral can therefore be regarded as limit of the sum

$$\sum_k F(x_k, y_k, z_k) \cos \gamma_k \cdot \sigma_k = \sum_k \frac{F(x_k, y_k, z_k)}{\sqrt{1 + f'_x{}^2(x_k, y_k) + f'_y{}^2(x_k, y_k)}} \sigma_k$$

when the division becomes indefinitely fine. But limit of the sums of the form

$$\sum_k \varphi(x_k, y_k, z_k) \sigma_k,$$

where σ_k are areas of the cells into which the part S is divided and (x_k, y_k, z_k) is an arbitrary point in the cell σ_k , as considered in § 120, example 2, where we have agreed to call this limit as integral of the function $\varphi(x, y, z)$ over the part S of the given surface and denote it by

$$\iint_S \varphi(x, y, z) d\sigma.$$

We therefore have in our case:

$$\begin{aligned} \iint_S F(x, y, z) dx dy &= \iint_S F(x, y, z) \cos \gamma(x, y, z) d\sigma = \\ &= \iint_S \frac{F(x, y, z) d\sigma}{\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}}. \end{aligned} \quad (2)$$

This formula is analogous to formula (6) § 121 and plays the same role for surface integrals as formula (6) § 121 plays in the theory of curvilinear integrals. Even in the simplest case which we are now considering it defines a surface integral constructively (since the right-hand side of formula (2) is obtained as the result of a construction and defined as limit of a sum of definite form). Further, as we shall soon see, formula (2), like formula (6) § 121, serves as the basis for an extension of the concept of surface integrals beyond the limits of the simplest case which we are considering in this paragraph.

It is self-evident that similar definitions and relations hold if we integrate with respect to the pair of variables (x, z) or (y, z) , provided the chosen part of the surface can be expressed by equations analogous to the equation (1). In practice one frequently meets sums of three integrals taken over the same part S which involve different pairs of variables.

Let $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ be three continuous functions in a given region in a three-dimensional space which contains the part S within itself. Let $\alpha = \alpha(x, y, z)$, $\beta = \beta(x, y, z)$, $\gamma = \gamma(x, y, z)$ be the angles between the normal to the given surface at the point (x, y, z) on the part S and the positive direction of the OX , OY and OZ axes respectively and let the direction of this normal be so chosen that the angles α , β and γ are acute on the part S . It is customary to write the sum of the integrals

$$\begin{aligned} \iint_S P(x, y, z) dy dz + \iint_S Q(x, y, z) dz dx + \\ + \iint_S R(x, y, z) dx dy \end{aligned}$$

in the form of one integral

$$\iint_S (P dy dz + Q dz dx + R dx dy);$$

if the chosen part S is such that each of the three coordinates on this part is a single-valued function of the other two and has continuous partial derivatives so that on the basis of the above discussion we evidently have :

$$\begin{aligned} \iint_S P dy dz &= \iint_S P \cos \alpha d\sigma, \quad \iint_S Q dz dx = \iint_S Q \cos \beta d\sigma, \\ \iint_S R dx dy &= \iint_S R \cos \gamma d\sigma, \end{aligned}$$

and consequently

$$\begin{aligned} \iint_S (P \, dy \, dz + Q \, dz \, dx + R \, dx \, dy) &= \\ &= \iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) \, d\sigma. \end{aligned}$$

§ 127. General definition of surface integrals

We have found at the beginning of § 126 that the simple definition of the surface integral

$$\iint_S F(x, y, z) \, dx \, dy \tag{1}$$

essentially implies that the given surface can be represented over the part S by an expression of the form (1) § 126, *i.e.* every straight line parallel to the OZ -axis intersects the part S in no more than one point. This condition is very restrictive in practice where we often integrate round closed surfaces which do not satisfy the above condition. We must therefore develop the concept of a surface integral beyond that considered in § 126 so that it covers more complicated cases occurring in practice.

A useful starting point for this development is given by formula (2) § 126. The middle part of this formula is, by definition, independent of any specific shape of the part S and remains fully valid, for example, for relatively simple closed surfaces. Although in the simplest case the relation (2) § 126 is proved as a *theorem*, we can nevertheless regard the first equation of formula (2) § 126 as the *definition* of the integral standing on its left-hand side. We are thus able to define the integral (1) for every part S on which the integral in the middle part of formula (2) § 126 exists. This gives us a much wider scope quite sufficient for many cases.

However, in order that this definition should be unique it is necessary even for surfaces of a more general form to indicate exactly the value of the angle $\gamma(x, y, z)$ in the middle part of formula (2) § 126; assuming that γ always denotes the angle between the normal to the given surface and the direction of the OZ -axis, we must establish a definite *direction for the normal* at every point on the given surface (which is, of course, equivalent to the choice of the sign of $\cos \gamma$). In the simplest case we have agreed always to direct the normal in the positive direction of the OZ axis, *i.e.* to make the angle

γ acute ($\cos \gamma > 0$). This agreement appeared natural at that time (in fact, it is the only possible one), since $\cos \gamma$ has first appeared in our arguments as the ratio of two areas. However, it can be readily seen that for figures of more general form this ratio would be inconvenient.

Let us imagine, for example, that S is the surface of a sphere. Then on the upper hemisphere we must take exterior points and for the lower hemisphere interior points on the sphere for the given direction of the normal; by passing through the equator the chosen direction would suddenly be reversed (fig. 93). It can be readily seen that a definition connected with this choice of direction of the normal will in most cases contradict the essence of the problem and unnecessarily complicate its solution. We shall therefore now introduce a new idea for the direction of the normal which is free from this disadvantage.

In order to make this new definition more concrete let us first return to the case when S is the surface of a sphere. It is immediately clear and also confirmed historically that in many practical cases connected with integrals of the form (2) § 126, it would be most convenient to direct the normal to every point either on the exterior or the interior of the sphere, *i.e.* to deal everywhere with an outward or inward normal. Let us select, say, the outward normal, *i.e.* let us agree that γ in the integral (2) § 126 denotes the angle between the outward normal to the sphere S at the point (x, y, z) and the positive direction of the OZ -axis. In that case γ will be acute for points on the upper hemisphere S_1 , be a right angle for points on the equator and an obtuse angle for points on the lower hemisphere S_2 . Therefore if $\iint_{S_1} F dx dy$ and $\iint_{S_2} F dx dy$ respectively denote surface integrals over the "simple" parts S_1 and S_2 defined in the sense given in § 126, then, in accordance with our new definition of the angle γ , we have

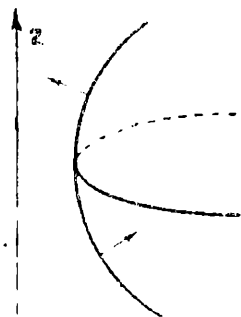


Fig. 93

$$\iint_{S_1} F \cos \gamma d\sigma = \iint_{S_1} F dx dy,$$

$$\iint_{S_2} F \cos \gamma d\sigma = - \iint_{S_2} F dx dy,$$

so that it follows from the extended definition of the surface integral (formula (2) § 126) that

$$\begin{aligned} \iint_S F dx dy &= \iint_S F \cos \gamma d\sigma = \iint_{S_1} F \cos \gamma d\sigma + \iint_{S_2} F \cos \gamma d\sigma = \\ &= \iint_{S_1} F dx dy - \iint_{S_2} F dx dy \end{aligned} \quad (2)$$

(where we must note that both integrals on the right-hand side are defined in the initial sense given in § 126). Our definition of the angle γ is such that the normal at every point on the sphere is directed outward so that $\cos \gamma > 0$ for points on the upper hemisphere and $\cos \gamma < 0$ for point on the lower hemisphere. In accordance with this definition we say about the integral (2) that it *extends over the exterior of the sphere* S . We could, of course, agree to take the inward normal instead of the outward normal at all points on the sphere S ; since $\cos \gamma$ changes its sign in the transition from the outward to the inward normal at every point of the surface S , therefore our two integrals will only differ from one another by their sign.

The position is even simpler when the part S has the shape considered in § 126, *i.e.* when it can be expressed by an equation of the form (1) § 126. We have selected the normal directed upwards ($\cos \gamma > 0$) and proved that

$$\iint_S F dx dy = \iint_S F \cos \gamma d\sigma.$$

In accordance with our new point of view we can therefore say that this last integral extends over the *upper part of* S . On the other hand, if we agree to direct the normal downwards, then the integral

$$\iint_S F \cos \gamma d\sigma \quad (3)$$

will reverse its sign; this will be an integral extending over the *lower part of* S . We thus see that the notation (3) does not tell us anything about the direction of the surface S over which we integrate; it must be stressed separately in every case.

Let us now consider the general case. Will S be a closed surface (like a sphere) or have a definite contour (boundary)? We

have seen in the last two examples that we can usually distinguish between two “sides” of this surface. If we are given a definite side of a surface, we are simultaneously also given a definite direction for the normal.

If a given side of the surface is chosen, then the *surface integral*

$$\iint_S F dx dy,$$

extending over this side of the surface is, by definition, equal to the integral

$$\iint_S F \cos \gamma d\sigma,$$

where γ is the angle between the chosen direction of the normal and the positive direction of the OZ -axis.

However, if we want our definition to cover the widest possible class of surfaces, we must consider in somewhat greater detail what is meant by the two “sides” of a surface mentioned above; even in elementary cases this problem may cause some difficulties.

Let us assume that the surface S in which we are interested (or part of this surface) has a tangential plane at every point whose direction changes continuously in relation to the continuous displacement of the point over the surface. We have said above that in order to choose a definite *side* of our surface it is sufficient to choose a definite direction for the normal at every point. However, if we choose this direction at different points of the surface S independent of each other, then, in general, we shall obtain nothing useful, for in this case the angle γ may everywhere be a discontinuous function of the position of the point so that integrals containing $\cos \gamma$ will be devoid of meaning. These formal considerations as well as visual representation clearly show that we can only choose the normal at every point on the surface so that its direction changes continuously as the point moves continuously over the surface; in other words, it is imperative that the angle γ should be a continuous function of the coordinates of the point to which it refers.

Let us now take an arbitrary point A on the surface S and draw a normal to the surface at this point and choose one of the two possible directions along this normal. We shall displace the point

A continuously over the surface and at every point through which we are passing we shall ascribe that possible direction to the normal to the given surface which leads us to it by following a continuous path. Let us assume that having covered some distance in this way we return to the point A . What will be the direction of the normal at this point on arrival — will it be the same as the initial direction or will it be the reverse? It can be readily seen that, as in the two examples considered above as well as in many other examples which we might think of, we shall always return to the point A so that the normal has the same direction as that with which we began (it is assumed that our path does not intersect at the edge of the surface S). However, this rule does not apply to all surfaces; let us consider, for example, the well-known “Moebius ribbon”, whose model can be obtained by cutting in paper the rectangle $abcd$ (fig. 94) and, after turning it, glueing the side ad to the side bc so that a coincides with c and d with b . It can be readily shown that having selected an arbitrary point A on this ribbon and having described the whole ribbon as explained above we shall on our return to A have reversed the direction of the normal. In future we shall almost entirely disregard such surfaces (which are exceedingly rare in practice) and assume that irrespective of the chosen starting point and the continuous path over the surface by which we may travel as long as it does not intersect the edge of the surface, the normal on our return to the starting point will have the same direction, provided it has changed continuously in our movement over the surface.

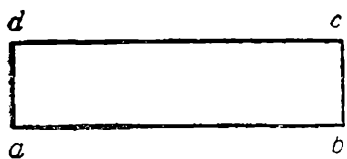


Fig. 94

This shows that it is sufficient to choose the direction of the normal at any given point A on the surface S for it to remain defined uniquely at every other point on this surface. In fact, if B is another arbitrary point on the surface, then two arbitrary path L_1 and L_2 from the point A to the point B will give us the same direction for the normal at the point B , for otherwise by starting from the point B and following the path L_1 in the opposite direction (to the point A) and then the path L_2 in the straight direction we should end at the point B with a normal whose direction would be opposite to that with which we started; and this, in accordance with our assumption, is impossible.

Hence for surfaces of the type considered above the choice of

direction of the normal at one point on the surface defines uniquely its direction at all other points on this surface, *i.e.* it defines uniquely the *side* of the surface; by choosing the opposite direction for the normal at the given point we simultaneously change its direction at all other points of the given surface and thus change to the other side of this surface. Hence surfaces of this type are known as *two-sided* (in contrast to *one-sided*, an example of which is the Moebius ribbon).

After these explanations we can return to the definition of an integral taken over the given side of the surface. Let D be a two-sided surface (or a part of such a surface) where we have chosen one side. Let $\gamma = \gamma(x, y, z)$ be the angle between the positive direction of the OZ -axis and that direction of the normal to the surface S at the point (x, y, z) which defines the side of the surface chosen by us; in accordance with our definition we can then assume that

$$\iint_S F dx dy = \iint_S F \cos \gamma d\sigma,$$

where we assume that the integral on the right-hand side exists and is defined as limit of sums of definite form similar to those described in § 126. This notation tells us nothing about the side chosen on the surface S and this fact must therefore be stressed separately in each case. If the surface S is closed, it is usual (and we shall do so in future) to consider that the above integral applies to the exterior of this surface so that γ is the angle between the outward normal to the surface S and the positive direction of the OZ -axis.

It is interesting to compare this wider definition of a surface integral with the initial definition for the simplest case given in § 126. Let the surface S (as is usual in many real cases) be divided into a finite number of "simple" parts S_1, S_2, \dots, S_n , etc. each of which can be expressed by an equation of the type (1) § 126. In accordance with our wider definition the integral taken over the given side of the surface S is equal to the sum of the integrals taken over the same side of its component parts so that evidently

$$\iint_S F \cos \gamma d\sigma = \sum_{i=1}^n \iint_{S_i} F \cos \gamma d\sigma.$$

However, on every part S_i the integral taken over the chosen side of the surface coincides with the integral defined in § 126, provided

$\cos \gamma > 0$, *i.e.* provided the chosen side of the surface S is the *upper* side; otherwise the absolute value of these two integrals would be the same but their signs opposite. Hence the integral taken over the given side of the surface is equal to the *algebraic sum* of the integrals of its component simple parts if these integrals are defined as in § 126; in this case integrals taken over upper sides of the surface S have the sign “+” and those taken over inner sides have the sign “-”.

Finally all that was said above in connection with integrals of the variables (x, y) evidently remains valid for integrals of the pairs (z, x) and (y, z) . For integrals taken over a definite side of a two-sided surface S we obtain the following general formula:

$$\begin{aligned} \iint_S (P \, dy \, dz + Q \, dz \, dx + R \, dx \, dy) = \\ = \iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) d\sigma, \end{aligned} \quad (4)$$

where P, Q, R are continuous functions of x, y, z in a region in space which contains the surface S within itself and α, β , and γ are the respective angles between the positive direction of the OX, OY , and OZ axes and that direction of the normal to the surface S at the point (x, y, z) which defines the chosen side of this surface,

In the remaining part of this chapter we shall assume that all surface integrals apply over a definite side of the given surface in accordance with the established definition.

Let us now make one final remark which is intended to clarify possible misunderstandings. In § 120 we have defined an *integral*

$$\iint_S \varphi(x, y, z) \, d\sigma \quad (5)$$

taken over a given part S of the surface. We have pointed out the analogy existing between this integral and the usual double integral taken over a part of a plane; in order to construct the integral (5) we must, as before, divide the part S into cells and multiply the area of each cell by a function φ at an arbitrary point in the given cell; we then and limit of the sum of all these products. Hence the integral

(5) results from the well-known construction which is typical for all problems of integral calculus.

In this paragraph we have defined the concept of a *surface integral*

$$\iint_S F(x, y, z) dx dy, \quad (6)$$

taken over a definite side of the given surface S . What is the relationship between these two types of integrals? This question can be clearly answered on the basis of the above arguments. The symbols (5) and (6) have a different meaning; the integral (5) only depends on the part S and form of the function φ but is quite independent of the choice of the "side" of this surface whereas the integral (6) changes its sign when the side chosen on the given surface is changed; only after a definite side is chosen, the integral (6) has a definite meaning; we can therefore say that the symbol (6) defines two separate integrals in relation to whether one or other side of the surface S is chosen.

For exercises to § 127, cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 403, 405.

§ 128. Ostrogradskij's formula

In § 123 we have deduced Green's formula which is very important in theory and practice; this formula connects a double integral taken over a plane region with a curvilinear integral taken over the contour of this region. A corresponding and no less important formula covering three-dimensional space was first deduced by M. V. Ostrogradskij; this formula connects a triple integral taken over a region in three-dimensional space with a surface integral taken over the exterior of the boundary of this surface.

Let us assume that we are given the region V in three-dimensional space bounded by a closed surface S . Let us assume at first that the shape of this surface is the simplest possible shape for a closed surface; every straight line parallel to one of the axes of co-ordinates intersects it at no more than two points so that S is divided into two parts — the "upper" and "lower" parts which can respectively be expressed by the equations $z = f_1(x, y)$ and $z = f_2(x, y)$; we shall assume that the functions f_1 and f_2 are continuous and have continuous partial derivatives with respect to x and y . Finally, let the function $R(x, y, z)$ and its partial derivative $\partial R / \partial z$ be defined and continuous in a region in space which

contains the region V within itself. Let us consider the triple integral

$$\iiint_V \frac{\partial R}{\partial z} dx dy dz.$$

According to formula (2) § 119 we can represent it in the form

$$\iint_D \left\{ \int_{f_2(x,y)}^{f_1(x,y)} \frac{\partial R}{\partial z} dz \right\} dx dy,$$

where D denotes the projection of the region V on to the XOY -plane. But

$$\int_{f_2(x,y)}^{f_1(x,y)} \frac{\partial R}{\partial z} dz = R[x, y, f_1(x, y)] - R[x, y, f_2(x, y)];$$

therefore we obtain

$$\begin{aligned} \iiint_V \frac{\partial R}{\partial z} dx dy dz &= \\ &= \iint_D R[x, y, f_1(x, y)] dx dy - \iint_D R[x, y, f_2(x, y)] dx dy. \end{aligned}$$

Let us consider the first of these two integrals. It follows from the definition of a surface integral that the first integral is an integral of the function $R(x, y, z)$ taken over the upper side $S_1 [z = f_1(x, y)]$ of the surface S and hence also over the *upper side* of the surface S . Similarly the second integral on the right-hand side is an integral of the function $R(x, y, z)$ taken over the upper side of the lower part $S_2 [z = f_2(x, y)]$ of the surface S ; but we know that in this case the same integral with its sign reversed (and in our formula it happens to have the “—” sign) will be the integral of the function R taken over the *lower side* of the surface S_2 which again coincides with the exterior of the surface S . We therefore obtain :

$$\iiint_V \frac{\partial R}{\partial z} dx dy dz = \iint_{S_1} R dx dy + \iint_{S_2} R dx dy,$$

where the first integral on the right-hand side is taken over the *upper* side of the surface S_1 and the second over the *lower* side of the surface S_2 ; since in both cases we integrate over the exterior of the surface S , the sum of the integrals on the right-hand side can be

replaced by one integral taken over the exterior of the whole surface S and we obtain the simple relation

$$\iiint_V \frac{\partial R}{\partial z} dx dy dz = \iint_S R dx dy = \iint_S R \cos \gamma d\sigma; \quad (1)$$

the second and third terms of these equations are integrals over the *exterior* of the surface S . We have deduced this relation for closed surfaces S which are intersected at no more than two points by any straight line parallel to the OZ -axis; it can, however, be shown that it also remains valid in much wider cases. At first we note that the relation (1) remains valid when the surface S consists not only of S_1 and S_2 but also of the cylindrical part S^* whose generating lines are parallel to the OZ -axis (fig. 95); in fact, at points of the part S^* the normal to the surface evidently makes a right-angle with the OZ -axis, $\cos \gamma = 0$, and we have :

$$\iint_{S^*} R dx dy = \iint_{S^*} R \cos \gamma d\sigma = 0,$$

so that as before

$$\begin{aligned} \iint_{S_1} R dx dy + \iint_{S_2} R dx dy &= \\ &= \iint_{S_1} R dx dy + \iint_{S_2} R dx dy + \iint_{S^*} R dx dy = \iint_S R dx dy, \end{aligned}$$

and formula (1) therefore remains valid.

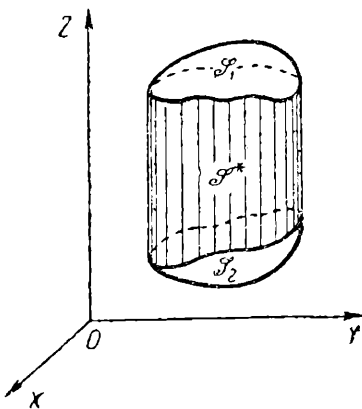


Fig. 95

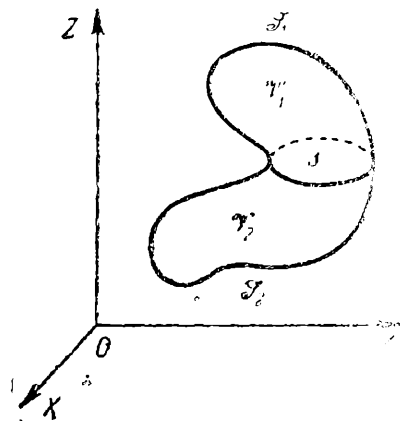


Fig. 96

We can also show that if the region V can be divided by means of a surface drawn within it into two parts V_1 and V_2 to each of which the relation (1) applies, then this relation also applies to the

whole of the region V . In fact, we have :

$$\begin{aligned} \iiint_V \frac{\partial R}{\partial z} dx dy dz &= \iiint_{V_1} \frac{\partial R}{\partial z} dx dy dz + \iiint_{V_2} \frac{\partial R}{\partial z} dx dy dz = \\ &= \iint_{S_1} R dx dy + \iint_{S_2} R dx dy, \end{aligned}$$

where S_1 and S_2 respectively denote closed surfaces which are the boundaries of the regions V_1 and V_2 and where each of the last two integrals is taken over the exterior of the corresponding surface. But S_1 is part of the surface S and the demarcating surface S (fig. 96) while S_2 consists of the remaining part of the surface S and the same demarcating surface S ; hence the latter sum of integrals can be represented as the sum of integrals over the exterior of the surface S and two integrals over the demarcating surface S ; of these two integrals one is taken over the side of the surface S which is the exterior for the surface S_1 whereas the other integral is taken over the side which is the exterior of the surface S_2 ; it is evident that these are two opposite sides of the surface S , as a result of which the sum of the two integrals vanishes and we obtain :

$$\iiint_V \frac{\partial R}{\partial z} dx dy dz = \iint_S R dx dy,$$

which we wanted to establish.

With the help of this theorem we can considerably widen the applications of formula (1) since most regions in space which are met with in practical cases can be divided by demarcating surfaces into regions of simple form for which formula (1) was initially deduced.

It is evident that all that is said above also refers to cases when the pair of the variables (x, y) is replaced by the pair of variables (y, z) or (z, x) . If $P = P(x, y, z)$, $Q = Q(x, y, z)$, $R = R(x, y, z)$ are continuous functions with continuous partial derivatives in a region in three-dimensional space which contains the region V bounded by the surface S within itself, then we obtain for a wide class of such regions :

$$\begin{aligned} \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx dy dz &= \\ &= \iint_S (P dy dz + Q dz dx + R dx dy), \quad (2) \end{aligned}$$

where the integral on the right-hand side is taken on the exterior of the surface S . We can evidently write this relation in the form

$$\begin{aligned} \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx dy dz = \\ = \iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) d\sigma, \end{aligned} \quad (3)$$

where α , β and γ respectively denote the angles between the outward normal to the surface S at the point (x, y, z) and the OX , OY and OZ axes.

Formula (2) (or formula (3)) is the above mentioned Ostrogradskij's formula ^{*)}. It is very important in many branches of physics, for it is the basis of the field theory; we shall return to this problem later. We shall now use Ostrogradskij's formula in order to prove a theorem which is important in connection with many problems of physics; this theorem is analogous to the corresponding proposition proved in § 124 by means of Green's formula.

Theorem. *In order that the integral*

$$\iint_S (P \cos \alpha + Q \cos \beta + R \cos \gamma) d\sigma,$$

*taken over any ^{**)} closed surface S which lies within the region V should be equal to zero it is necessary and sufficient that the following relation should be satisfied at every interior point of the region V :*

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = 0. \quad (4)$$

Proof. The sufficiency of the condition (4) can be seen directly from formula (3). In order to prove its necessity let us assume that at an interior point A of the region V , for example,

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} > 0.$$

It follows from the assumed continuity of the partial derivatives that this inequality will also be satisfied inside and on the boundary of a

^{*)} This formula is sometimes also known as the Gauss formula or the Gauss-Ostrogradskij's formula.

^{**)} We naturally have in mind a surface to which Ostrogradskij's formula can be applied; this theorem remains valid even when the class of surfaces S is made much narrower, for example, when we are only considering spherical surfaces.

certain sphere v with centre at A which lies entirely within the region V ; hence

$$\iiint_v \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx dy dz > 0.$$

But formula (3) then shows that for the surface s of the sphere v

$$\iint_s (P \cos \alpha + Q \cos \beta + R \cos \gamma) d\sigma > 0.$$

This fully proves our proposition.

For exercises to § 128 cf. Problem Book by B. P. Demidovich, Section VIII, Nos. 415, 416, 423, 424.

§ 129. Stoke's formula

Ostrogradskij's formula introduced in the previous paragraph can be regarded as an analogue to Green's formula in transition from a plane to the three-dimensional space. We shall now deduce another equally important formula which involves the direct *generalisation* of Green's formula when a plane figure is replaced by a part of a curved surface. This formula connects an integral taken over a definite side of a part S of the surface bounded by the closed contour L with a curvilinear integral in space taken over its contour, *i.e.* it solves the same problem for a curved surface as is solved by Green's formula for a plane.

Let us first assume that the part S of the given surface is expressed by the equation $z = f(x, y)$; we shall make the usual assumptions as to continuity and differentiability with regard to the function f and the functions P, Q, R which we shall introduce below. Let us begin by considering the curvilinear integral in space

$$\int_L P(x, y, z) dx, \quad (1)$$

taken round the contour L of the part S and described in the direct way (*i.e.* so that to an observer moving over the upper side of the part S this part should always remain to the left of his path). Let the plane region s with contour denoted by l represent the projection of the part S on to the XOY -plane; in that case the integral (1) can also be represented by a curvilinear integral in a plane

$$\int_l P[x, y, f(x, y)] dx, \quad (2)$$

taken over the contour l ; in fact, let l' be a part of the contour l expressed by the equation

$$y = \varphi(x) \quad (a \leq x \leq b);$$

where l' is the orthogonal projection on to the XOY -plane of a part L' of the contour L which is evidently expressed by the equations

$$y = \varphi(x), \quad z = f[x, \varphi(x)],$$

and, in accordance with the initial definition of a curvilinear integral we have:

$$\begin{aligned} \int_{L'} P(x, y, z) dx &= \int_a^b P\{x, \varphi(x), f[x, \varphi(x)]\} dx, \\ \int_{l'} P[x, y, f(x, y)] dx &= \int_a^b P\{x, \varphi(x), f[x, \varphi(x)]\} dx; \end{aligned}$$

the right-hand sides, and therefore also the left-hand sides, coincide, which proves our proposition for simple sections l' and L' . Assuming that the contour L (and therefore also l) can be divided into a finite number of such simple sections we can directly see that in this case

$$\int_L P(x, y, z) dx = \int_l P[x, y, f(x, y)] dx; \quad (3)$$

this equation also holds in more general cases but we shall not stop here to prove it.

On the other hand, let us consider the integrals

$$\left. \begin{aligned} \int_S \frac{\partial P}{\partial y} dx dy &= \int_S \frac{\partial P}{\partial y} \cos \gamma d\sigma, \\ \int_S \frac{\partial P}{\partial z} dx dz &= \int_S \frac{\partial P}{\partial z} \cos \beta d\sigma, \end{aligned} \right\} \quad (4)$$

taken over the upper side of the surface S , where the angle γ and β are defined as usual. It follows from the formulae of § 99 that

$$\begin{aligned} \cos \beta &= \frac{\frac{\partial f}{\partial y}}{\pm \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}}, \\ \cos \gamma &= \frac{1}{\pm \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}}, \end{aligned}$$

where the sign of the denominator should be the same in both cases; since we have chosen the upper side of the surface S , therefore $\cos \gamma > 0$ and the positive radical must be taken; regardless of this we also have:

$$\cos \beta = -\frac{\partial f}{\partial y} \cos \gamma,$$

as a result of which the formula (4) give:

$$\begin{aligned} \iint_S \left(\frac{\partial P}{\partial y} dx dy - \frac{\partial P}{\partial z} dx dz \right) &= \iint_S \left(\frac{\partial P}{\partial y} + \frac{\partial P}{\partial z} \frac{\partial f}{\partial y} \right) \cos \gamma d\sigma = \\ &= \iint_S \left(\frac{\partial P}{\partial y} + \frac{\partial P}{\partial z} \frac{\partial f}{\partial y} \right) dx dy. \end{aligned}$$

If we replace z by $f(x, y)$, the integrand in this last integral will evidently be equal to

$$\frac{\partial}{\partial y} \{P[x, y, f(x, y)]\};$$

it therefore follows from the definition of a surface integral for the simplest case (§ 126) that this last integral can be written in the form of a double integral

$$\iint_S \frac{\partial}{\partial y} \{P[x, y, f(x, y)]\} dx dy,$$

and we obtain:

$$\iint_S \left(\frac{\partial P}{\partial y} dx dy - \frac{\partial P}{\partial z} dx dz \right) = \iint_S \frac{\partial}{\partial y} \{P[x, y, f(x, y)]\} dx dy. \quad (5)$$

Comparing the formulae (3) and (5) we can see directly that, as a result of Green's formula (§ 123), the right-hand sides only differ from one another by their sign. Hence the same also applies to the left-hand sides and we obtain:

$$\iint_S \left(\frac{\partial P}{\partial z} dx dz - \frac{\partial P}{\partial y} dx dy \right) = \int_L P(x, y, z) dx. \quad (6)$$

Here the integral on the left-hand side is taken over the upper side of the surface S and the integral on the right-hand side is in such a direction that an observer moving over the upper side of S should have the part S on his left. If we change the chosen side S and the direction in which we describe the contour L , then both sides of the

equation (6) will change their sign so that this equation will remain valid ; it can be readily seen that in this case also an observer standing on the chosen (in this case lower) side of the surface and moving along the contour L in the new changed direction will have the part S on his left. Hence this rule is of quite general character and defines uniquely the direction of movement over the contour L provided a definite direction is chosen for the surface S (and vice versa).

Formula (6), like Green's and Ostrogradskij's formulae, possesses the property that if the part S is divided by means of a line drawn on it into two parts S_1 and S_2 to each of which this formula applies, then this formula also remains valid for the whole parts S (the proof is exactly the same as for Green's formula and we can leave it to the reader). With the help of this formula the validity of formula (6) (as for Green's and Ostrogradskij's formulae) can be established for a wide class of surfaces S .

The circular transposition of the letters x, y, z on the one hand and P, Q, R on the other gives us two more formulae besides formula (6) :

$$\iint_S \left(\frac{\partial Q}{\partial x} dy dz - \frac{\partial Q}{\partial z} dy dx \right) = \int_L Q(x, y, z) dx,$$

$$\iint_S \left(\frac{\partial R}{\partial y} dz dx - \frac{\partial R}{\partial x} dz dy \right) = \int_L R(x, y, z) dy.$$

Finally adding all three formulae we obtain the general *Stoke's formula*

$$\iint_S \left\{ \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dy dz + \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) dz dx + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) dx dy \right\} =$$

$$= \int_L (P dx + Q dy + R dz), \quad (7)$$

which we were trying to deduce.

We have already said that this formula is a generalisation of Green's formula : if S is a part of the XOY -plane, then Stoke's formula becomes Green's formula as can be readily seen ; hence the latter is a particular case of the former.

Here we shall only give one of the numerous applications of Stoke's formula. In § 125 we were unable to prove the following

lemma : if at every point of region V in a three-dimensional space the following equations hold :

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}, \quad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}, \quad (8)$$

then the curvilinear integral

$$\int_L (P dx + Q dy + R dz),$$

taken along an arbitrary closed curve L which lies entirely within the region V is equal to zero. The proof of this proposition follows directly from Stoke's formula provided (which we must assume) that for every closed curve L in the region V a surface S exists which is bounded by the curve L and to which Stoke's formula applies (in most practical cases these conditions are satisfied^{*)}). In fact, it follows from the condition (8) that on any closed curve L which lies entirely within the region V the left-hand side, and therefore also the right-hand side, of Stoke's formula is equal to zero which proves the above lemma.

§ 130. Elements of the field theory

Mechanical and physical applications of the theory of curvilinear and surface integrals have a common mathematical basis which is usually called the field theory. Some elements of this theory have already been considered in the last few chapters. However, in mechanics and physics a more visual terminology is usually preferred for quite understandable reasons in order to define the concepts and relations of the field theory; in practice these concepts and relations are usually formulated in the vector form; this makes them simpler and clearer and facilitates solution of the problems. We shall therefore briefly consider the more important concepts and relations in vector form and thus make them more accessible^{**)}.

1. *Scalar and vector fields.* Quantities involved in mechanics and physics can essentially be divided into two groups: *scalars*, i.e.

^{*)} We shall mention, however, a simple case when the required surface S does not exist. Let the curve L be a circle of unit radius and the region V a set of points in space at distances more than $\frac{1}{2}$ from the circle (the surface of such a region is known as a "torus"). In this case there is evidently no surface which wholly belongs to the region V and which is bounded by the circle L .

^{**)} We assume that the reader is familiar with the elements of vector algebra.

quantities which are fully characterised by their numerical values (density, temperature, electrical potential) and *vectors* whose full characteristic involves direction as well as numerical value (velocity, acceleration, force). In physics other quantities are also used but their definition is even more complicated; however, we shall not consider them here.

A *component* of a vector along an axis (a directed straight line) is, as we know, the projection of the vector onto this axis. We shall, in future, print vectors with bold-faced letters. The numerical value (non-negative) of the vector \mathbf{F} is usually denoted by $|\mathbf{F}|$ and its components along the axes of coordinates by F_x , F_y and F_z respectively.

If we have a scalar quantity F defined at every point in space or in a given part of it, then the set of these quantities F is known as the scalar field. The definition of a scalar field evidently does not differ in any way from the definition of a function $F(x, y, z)$ of the coordinates of a point. We are given a *vector field* if at every point in space (or in a part of it) the vector \mathbf{F} is defined (both in magnitude and direction); in order to do this it is sufficient to determine the three components F_x , F_y , F_z of the vector \mathbf{F} at every point. Hence the definition of a vector field is equivalent to defining three functions $F_x(x, y, z)$, $F_y(x, y, z)$ and $F_z(x, y, z)$ of the coordinates of the points in space.

2. *Level surfaces and gradient of a scalar field.* Let us assume that we are given a scalar field $F(x, y, z)$ in space or in a part of it; we shall always assume that the function F has continuous partial derivatives of the first order in the given part of space. The equation

$$F(x, y, z) = C,$$

where C is an arbitrary constant defines in general a surface which we shall call *level surface* of the given scalar field; it is quite clear that one and only one level surface of the given field can pass through every point in space so that two different level surfaces cannot have common points (they cannot intersect).

We know (§ 91) that the rate of change of a function $F(x, y, z)$ at the given point (x, y, z) in the given direction λ is measured (in its absolute value and sign) by the derivative $D_\lambda F$ of the function F in this direction; this derivative (§ 91) is expressed by the formula

$$D_\lambda F = \frac{\partial F}{\partial x} \cos \alpha + \frac{\partial F}{\partial y} \cos \beta + \frac{\partial F}{\partial z} \cos \gamma,$$

where α , β and γ denote the respective angles between the direction λ and the positive direction of the OX , OY and OZ axes.

Let us now consider the scalar field $F(x, y, z)$ together with the vector field $\mathbf{G}(x, y, z)$ which is defined by the relations

$$G_x = \frac{\partial F}{\partial x}, \quad G_y = \frac{\partial F}{\partial y}, \quad G_z = \frac{\partial F}{\partial z},$$

so that

$$|\mathbf{G}(x, y, z)| = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2 + \left(\frac{\partial F}{\partial z}\right)^2}.$$

The vector $\mathbf{G}(x, y, z)$ is called *gradient* of the scalar field F (at the given point (x, y, z)) and denoted $\text{grad } F$. In accordance with the known laws of vector algebra the quantity

$$D_\lambda F = G_x \cos \alpha + G_y \cos \beta + G_z \cos \gamma$$

is the projection of the vector \mathbf{G} in the direction λ , i.e.

$$D_\lambda F = |\mathbf{G}| \cos(\mathbf{G}, \lambda),$$

where (\mathbf{G}, λ) denotes the angle between the direction of the gradient \mathbf{G} and the given direction λ .

If we compare the rate of change $|D_\lambda F|$ of the function F in different directions λ at a given point (x, y, z) , then the last equation shows that this rate will reach its maximum if $|\cos(\mathbf{G}, \lambda)| = 1$ i.e. if the direction λ coincides with the direction of the gradient (or is opposite to it). Hence *the direction of the gradient at every point gives the direction of the maximum rate of change of the given scalar field; this maximum rate of change is expressed by*

$$|\mathbf{G}| = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2 + \left(\frac{\partial F}{\partial z}\right)^2}.$$

Let us finally note that since $\partial F / \partial x$, $\partial F / \partial y$, $\partial F / \partial z$ are proportional (§ 99) to the direction cosines of the normal to the surface $F(x, y, z) = C$ at the point (x, y, z) , *the direction of the gradient coincides with the direction of the normal to the level surface which passes through the given point.* The magnitude $|\mathbf{G}| = |\text{grad } F|$ of the gradient is measured by the absolute value of the derivative of the normal to the level surface; denoting this derivative by $\partial F / \partial n$ we have:

$$|\text{grad } F| = \left| \frac{\partial F}{\partial n} \right|.$$

3. *Divergence of a vector field and flow across a given surface.* Let us assume that we are given in space, or in a part of it, a vector field $\mathbf{F}(x, y, z)$. In practice the following quantity is very important:

$$\text{div } \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z},$$

which is known as *divergence* of the vector field \mathbf{F} at the point (x, y, z) . Divergence of the field at a given point is a scalar quantity and the set of its values in the part of space under consideration forms the scalar field.

We shall now define a concept which is very important in practical applications of mechanics and physics, *i.e.* the concept of *flow of a given vector field across the given surface*. In order to make this concept more accessible we shall illustrate it by a hydrodynamic model. Let us assume that the chosen part of space is filled with a moving fluid; let us take a two-sided surface S in a region in space which is either open or bounded by a closed contour and choose on this surface a definite side in the sense defined in § 127. Let $d\sigma$ be an element (a very small area) of the surface S . If at a given instance of time the rate of flow at a point on the element $d\sigma$ is expressed by the vector \mathbf{F} , then the quantity of fluid which flows across this element in a short time interval dt in the direction of the chosen side of the surface will evidently be equal (with an accuracy to infinitely small quantities of higher orders) to the quantity of fluid contained in a cylinder with a base $d\sigma$ and the height $|\mathbf{F}| dt$, whose generating lines are straight lines parallel to the vector \mathbf{F} . The volume of this cylinder is evidently equal to $|F_n| dt d\sigma$, where F_n is the projection of the vector \mathbf{F} in the direction of the normal to the surface $d\sigma$ corresponding to the chosen side of the surface. Hence the quantity of fluid which flows across the element $d\sigma$ during the time interval dt can be written in the form

$$\rho F_n dt d\sigma,$$

where ρ is density of the fluid and mass of the flowing fluid can be positive or negative according as the fluid flows in the chosen direction ($F_n > 0$) or in the opposite direction ($F_n < 0$). The mass of fluid which flows across the area $d\sigma$ in unit time is therefore equal to :

$$\rho F_n d\sigma,$$

and the mass of fluid which flows in unit time across the surface S can be expressed by the surface integral

$$M = \iint_S \rho F_n d\sigma.$$

If we denote by α, β, γ the angles between the chosen direction of the normal to the surface S and the positive direction of the axes of coordinates, then, in accordance with the laws of vector algebra, we have :

$$F_n = F_x \cos \alpha + F_y \cos \beta + F_z \cos \gamma$$

and the quantity M (if for the sake of simplicity we assume that ρ is constant) can be written in the form

$$M = \rho \iint_S (F_x \cos \alpha + F_y \cos \beta + F_z \cos \gamma) d\sigma.$$

For reasons which are now obvious the surface integral

$$\iint_S (F_x \cos \alpha + F_y \cos \beta + F_z \cos \gamma) d\sigma = \iint_S F_n d\sigma,$$

taken over a definite side of the surface S is known as *flow of a vector field $\mathbf{F}(x, y, z)$ across the surface S* in the direction defined by the chosen side of the surface. We have met integrals of this type on many occasions in the past. For example the surface integral in Ostrogradskij's formula (3) (§ 128) is an integral of this type if we assume (which is, of course, always possible) that P, Q and R are the components F_x, F_y, F_z of a vector \mathbf{F} in the direction of the coordinate axes; if we also pay attention to the fact that the integrand on the left-hand side of formula (3) § 128 represents in this case divergence of the vector \mathbf{F} , then Ostrogradskij's formula can be written in the form

$$\iiint_V \operatorname{div} \mathbf{F} dx dy dz = \iint_S F_n d\sigma,$$

which is at the same time very simple and expressive. Hence this formula implies in the vector sense that flow of a vector field from the interior of a closed surface is equal to the integral of divergence of this field in a region bounded by the given surface. Also the theorem proved at the end of § 128 can be stated as follows: *in order that flow of the given vector field across an arbitrary closed surface in the given region V should be equal to zero it is necessary and sufficient that divergence of this field should be identically zero in this region V .*

4. *Circulation in a vector field. Vector of turbulence. Potential field.* We will now show that Stoke's formula (§ 129) can be given a simple and convenient vector interpretation. On the right-hand side of this formula ((7) § 129) we have the integral

$$\int_L (P dx + Q dy + R dz),$$

which, in accordance with § 125, we can represent in the form

$$\int_L (P \cos a + Q \cos b + R \cos c) d\lambda,$$

where a, b, c are the angles between the tangent to the curve L and

the positive directions of the coordinate axes. If we consider the vector field $\mathbf{F}(x, y, z)$ with the components

$$F_x = P, F_y = Q, F_z = R, \quad (1)$$

then the integrand evidently becomes the projection F_l of the vector \mathbf{F} in the direction of the tangent to the curve L at the given point and our integral can be written in the form

$$\int_L F_l d\lambda.$$

This integral is usually known as *circulation* of the vector field \mathbf{F} around a closed path L ; it is evident that F_l represents the projection of the vector \mathbf{F} onto the tangent to the curve L in the direction in which this contour is described.

Let us now consider the left-hand side formula (7) § 129. We know from formula (4) § 127 that this left-hand side can be written in the form

$$\iint_S \left\{ \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \cos \alpha + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \cos \beta + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \cos \gamma \right\} d\sigma,$$

where the angles α , β and γ are defined as usual (the direction of the normal must, of course, correspond to the chosen side of the surface S which, in its turn, must coincide with the direction in which the contour L on the right-hand side of this formula is described). Together with the vector field $\mathbf{F}(x, y, z)$ with components (1) let us now introduce another uniquely defined vector field $\mathbf{C}(x, y, z)$ whose component functions are

$$C_x = \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}, C_y = \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}, C_z = \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y};$$

the vector \mathbf{C} , which is very important in hydrodynamics, is known as *vector of turbulence* or *rotor* of the given field \mathbf{F} and the set of its values as *field of turbulence* (with respect to the field \mathbf{F}). The vector of turbulence \mathbf{C} is frequently denoted by $\text{rot } \mathbf{F}$ ("rotor of the vector \mathbf{F} "). Hence the last integral can be written in the form

$$\iint_S (C_x \cos \alpha + C_y \cos \beta + C_z \cos \gamma) d\sigma = \iint_S C_n d\sigma,$$

where C_n is the projection of the vector of turbulence $\mathbf{C} = \text{rot } \mathbf{F}$ in the direction of the normal to the surface S corresponding to the chosen side of the surface S . Hence Stoke's formula can be written as follows:

$$\iint_S C_n d\sigma = \int_L F_l d\lambda,$$

where the chosen side of the surface S and the direction in which the contour L is described coincide in accordance with § 129. Hence the vector interpretation of this formula involves the fact that *flow of a turbulent field \mathbf{C} across the given surface S bounded by the contour L is equal to the circulation of the given vector field \mathbf{F} along this contour* (where the direction of flow and the direction in which the contour L is described coincide).

To give an example of the general theorems of the field theory we shall state in vector terminology the lemma which we have proved at the end of § 129 by means of Stoke's formula; *if in a certain region V in space the vector field \mathbf{F} is such that the corresponding field of turbulence does not exist (i.e. $\text{rot } \mathbf{F} = 0$ at every point of the region V), then the circulation of the field \mathbf{F} becomes zero around every closed curve which can serve as the contour of a surface situated entirely within the region V .* The converse theorem is also valid as can be readily seen from the lemmas stated at the end of § 125.

If for the given vector field \mathbf{F} the vector of turbulence $\text{rot } \mathbf{F}$ vanishes at every point of a region V in space, then this shows that in this region

$$\frac{\partial F_x}{\partial y} = \frac{\partial F_y}{\partial x}, \quad \frac{\partial F_x}{\partial z} = \frac{\partial F_z}{\partial x}, \quad \frac{\partial F_y}{\partial z} = \frac{\partial F_z}{\partial y}$$

but it follows from the lemmas § 125 that these conditions are necessary and sufficient in order that the expression

$$F_x dx + F_y dy + F_z dz$$

should be differential of a function $U(x, y, z)$, i.e. a function U should exist for which

$$\frac{\partial U}{\partial x} = F_x, \quad \frac{\partial U}{\partial y} = F_y, \quad \frac{\partial U}{\partial z} = F_z. \quad (2)$$

If this function exists, it is known as *potential* or *potential function* of the field \mathbf{F} and the field \mathbf{F} itself as *potential field*. Finally if the relations (2) are satisfied, the vector \mathbf{F} evidently represents gradient of the scalar field defined by the function U . It therefore follows from above that *potential field, non-turbulent field* (i.e. for which the vector of turbulence becomes identically zero) and *gradient field* ($\mathbf{F} = \text{grad } U$) are equivalent concepts (at least for regions of sufficiently simple form). Thus we have for any scalar field $U(x, y, z)$

$$\text{rot grad } U = 0.$$

For exercises to § 130 cf. Problem Book by B.P. Demidovich, Section VIII, Nos. 436, 438, 439, 452, 468, 483.

CONCLUSION

Short historical sketch

I

During the XVII century practical requirements connected with the development of social and economical relationships which, in their turn, were influenced by technical progress in all fields of human activity confronted mathematicians with many new problems. A group of problems mainly connected with geometry and mechanics which became different from many earlier problems and required entirely new methods for solution soon became prominent. Many leading thinkers of that period naturally directed their attention towards solution of these problems. Most of the scholars could not clearly tell the time when this division of new problems from the old took place. However, we can now see it very clearly: the new problems arose in connection with the study of quantities where attention was centred not on the values of these quantities at a given moment of the process but on their *character of change* in the given phenomenon.

As is usually the case in the development of new fields of mathematical sciences, the methods for solving these new problems were gradually forthcoming, generally as a result of investigation of individual concrete problems; however, their common characteristics which form the basis of this scientific branch were recognised and explained rather slowly and could only be fully realised after many concrete problems were solved. At present we can clearly see that many problems revolved around two centres which we now call *differentiation and integration of functions*. In both cases the concept of function was the basis of the newly created science, *i.e.* the concept of a quantity which changes in strict relation to changes in other quantities. Hence from the methodical point of view the requirements for this new science to progress strictly corresponded to the

dialectical principles of nature study : not the momentary values of quantities were studied but their process of change ; on the other hand, quantities were studied not in isolation but in strict interdependence. Engels mentioned quite correctly somewhat later that motion and dialectics entered mathematics together with the variable quantity and became the main object of investigation.

The main outlines of the new science, for whose development most of the outstanding scientists of XVII century worked, only became apparent about the end of the century and were expressed in the fundamental works by Newton and Leibnitz. These works are usually regarded as the origin of differential and integral calculus. Newton and Leibnitz, independent of each other and by rather different methods, developed, the basis of the new science by generalising earlier researches. The great value of the works of both scholars is rightly ascribed to the fact that they were the first to give deserved prominence to the relationship between integration and differentiation of functions (§ 50) which reveals the mutually reciprocal character of these two basic operations of mathematical analysis and thus from that moment onwards, become historically, the mainspring for further development of this scientific branch. They also introduced infinite series which soon became an important investigational tool of mathematical analysis.

II

From XVIII century onwards differential and integral calculus began to develop rapidly and were accompanied by the creation of other scientific branches within mathematics itself (differential equations, calculus of variations and then integral equations, general functional analysis, etc.) and the pronounced penetration of the methods of "analysis of infinitely small quantities" into the ever-widening circle of applied sciences. We can say without exaggeration that during its whole course of development mathematics never knew an era when so much was achieved in such a short time and when its outlines were radically changed and infinitely widened.

In the fields of differential and integral calculus mean-value theorems were proved which, together with the development of infinite series, made the expansion of functions into series, *i.e.* at first into power series possible ; thus the accurate investigation of these series became possible. The methods of integration soon became well-known, new transcendental functions created by the process of integration were subjected to systematic investigations and a series

of important functions were defined by means of integrals depending on parameters (see beginning of chapter XXVI). The laws of differential and integral calculus were progressively applied to the study of functions of several variables. It is impossible to list the names of all leading mathematicians of this period who participated in this development. However, the names of two great scholars who contributed during the first stages of this new era must be noted, *viz.* Euler and Lagrange—who pointed out many new directions which proved important in the subsequent development of analysis. The St. Petersburg scientist Euler is not only known as the author of a series of special studies (“Euler’s substitutions” § 63, “Euler’s integrals” § 112, theorem on homogeneous functions § 93, etc.) but also as one of the creators of the theory of differential equations and calculus of variations: he also widened and rationalised concept of infinite series introduced by Newton and was the first to define a most important concept, *viz.* analytic functions; the works of Euler contain a great variety of applied problems which he solved by means of the new methods. Lagrange discovered the fundamental theory of mean-value theorems and was first to use them systematically, for example in the evaluation of the remainder of Taylor’s series as well as in the general development of power series; they were also the first to describe the elements of calculus of variations as an independent branch of mathematical analysis by introducing the concept of variation and establishing rules for variations. However, the most outstanding contribution to the new science by Lagrange was the creation of “analytic mechanics” which involved systematic construction of theoretical mechanics by means of the methods of analysis of infinitely small quantities. His was the first systematic work in this field although he based it on Euler’s works; his work was characterised by such finality that it retains its fundamental importance to the present day in spite of the great subsequent developments in mechanics.

Apart from mechanics and sometimes along with it the methods of analysis of infinitely small quantities began to penetrate rapidly into other branches of mathematics (geometry) and an ever-widening circle of applied sciences. Among the applied sciences in which the use of the new methods was particularly fruitful were many branches of mathematical physics (theory of heat, acoustics, electro-dynamics, theory of diffusion and many others) and the mathematical theory of probability (Bernoulli, Moivre, Laplace). The use of methods of mathematical analysis in many branches of theoretical and applied

sciences continued throughout the XIX century; it continues even today when applications of differential and integral calculus are so wide that every engineer must be familiar with the principles of this science in order to satisfy his everyday requirements. By studying the history of the gradual conquests made by the methods of infinitely small quantities we can clearly see the reason behind this victorious development. It is undoubtedly due to the fact that this method satisfied the requirements of the dialectico-methodical theory of learning and created a mathematical apparatus which was able to embrace the main outlines of many phenomena in the outside world.

III

We have thus seen that during the XVIII century results achieved with the help of the method of infinitely small quantities represent an impressive and unprecedented picture of wealth; however, the position was quite different where the logical basis of this new science is concerned. The structure of differential and integral calculus and its manifold applications went forward so rapidly and with such success that no time was left for revision and improvement of the theory. And here the position was very unfavourable. The logical basis on which this course was constructed was mainly created during the XIX century; even the elementary theory of limits which we have introduced in chapter 2 and defined more accurately and developed in the subsequent two chapters in order to bring it up to modern levels—even this imperfect theory was quite unknown in the XVIII century.

The situation was very peculiar at times: no fundamental concept of analysis was defined with any degree of accuracy; the question of what was meant by an infinitely small quantity was the subject of endless discussions which were quite useless from a logical point of view since in most cases neither side was able to present anything but vague examples which led nowhere. The position was similar with the concept of continuity, differential, derivative and integral. Imagine the difficulty—to teach these concepts to a man who is unfamiliar with the concept of the limit—and you will immediately realise that you can give him nothing but descriptions which cannot even be correct*.

At present we regard the concept of derivative as the main-spring of differential calculus whereas differential is of secondary

*) We must note, however, that some leading scholars understood certain aspects of the new ideas in almost modern sense.

importance, for it is defined by the derivative. In the XVIII century, however, differential was regarded as fundamentally important (although in the works of Newton and even more so in the works of Euler and Lagrange, concepts are defined more closely to the modern point of view). What did these scholars understand by differential? If y is a continuous function of x , then as Δx decreases, Δy also becomes infinitely small and a definite value of Δy corresponds to each value of Δx . We must imagine that at the very last moment before Δx and Δy vanish, these quantities assume their "last" values which are less than any of their previous values, but which are nevertheless not zero. These values were defined as infinitely small increments or as differentials of x and y and denoted by dx and dy ; their ratio $y' = dy/dx$ (we can divide since dx is not yet equal to zero!) was known as derivative of y with respect to x . Hence, differentials were regarded not as variables but as constants and derivative, not as a limit of the ratio of variable increments but as a real ratio of two increments. Therefore, the general concept of an infinitely small quantity assumed the same characteristics: an infinitely small quantity was believed to be the last stage in the decrease of a given quantity, *i.e.* a value which is less than all other values, so that after it only zero follows, although it is itself non-zero, *i.e.* it is a constant which cannot decrease further; therefore, such quantities were also often called "indivisibles". Accordingly, an integral was regarded not a limit of sum of an indefinitely increasing number of indefinitely decreasing terms, but as the real sum of an indefinite number of such "indivisibles". For the same reason—the absence of an accurate concept of limiting processes—the sum of an infinite series was regarded as the result of a real addition of an infinite number of terms. It is obvious that this concept of summation of series, made it impossible to define convergence with any degree of accuracy; the irresponsible operation with series whose convergence could not be established, was one of the main faults of this mathematical era which frequently led to mistakes and paradoxical results.

We can now clearly see that these points of view could not be preserved, since every attempted logical formulation revealed logical contradictions. However, even mathematicians of the XVIII century, at least some of them, clearly understood the faults of the theoretical basis on which the new science was constructed, but, unlike us, they were unable to replace it by something better. Works were published from time to time in which the currently accepted logical fundamentals of mathematical analysis were subjected to severe

criticism and sometimes even to ridicule. However, the real victories of this new science were so numerous and so significant that all doubts of this kind were unable to stop the powerful creative efforts of the builders of the new structure, while revision and improvement of the theoretical basis was left for a later period. D' Alembert's slogan: "Go forward and you will be reassured later" is characteristic of the mood of these times.

IV

This assurance, in fact, came but not until the XIX century when most essential problems of Analysis and its main applications had time to ripen and be solved. It was now possible to slacken the pace and devote greater attention to the revision of fundamentals; on the other hand, the development of the new theory stimulated critical tendencies, as is frequently the case, and made the position existing in the logical aspects of analysis quite insufferable.

In the twenties of the XIX century works were published (first to appear was Cauchy's "Course of Analysis") in which mathematical analysis was constructed on the new basis, *i.e.* on the theory of limiting process in its modern sense. Concepts like infinitely small quantities, continuity, differential and integral were already clearly defined; the sum of an infinite series was now treated as limit of its partial sums and not as a result of addition of an infinite number of terms; hence, an accurate definition of convergence became possible and use of divergent series was strictly limited. The proof of existence of integrals was given and differential equations were solved for the first time. As usual in such cases reconstruction of fundamentals cannot be ascribed to Cauchy alone; the time was now ripe for these new ideas which formed in the right direction in many leading brains of this era. Several years before the publication of Cauchy's "Course", the Czechoslovakian philosopher-mathematician Bolzano obtained a series of results which anticipated many ideas found in Cauchy's book, and also contained modern definitions of continuity and the first example of a continuous function without derivatives. Simultaneously with Cauchy, Abel deduced fundamental results which led to the creation of a strict theory of infinite series. Nevertheless, Cauchy's "Course" is undoubtedly the first work of this kind where an extensive thesis on analysis of infinitely small quantities based on logical considerations is given; this book served for a long time as a model for other works devoted to the same subject. However, in the next decade, Cauchy's

conceptions had to be defined rather more accurately and some corrections had to be made; for example, Cauchy failed to define the concept of uniform convergence of a series; he proved the theorem (which we now know to be incorrect) that the sum of a convergent series of functions which are continuous in an interval is also always continuous in that interval. Also, the concept of limit whose definition remained in principle unchanged to this day had to be defined more exactly; we did this in § 14, § 15 and used it subsequently throughout this book.

The greatest event in further historical development of the logical basis of mathematical analysis after Cauchy's era can undoubtedly be ascribed to the advances in the general theory of real numbers which evolved in the seventies of the XIX century. The necessity for this theory was felt as acutely at this time as the necessity for a clear definition of infinitely small quantities was felt at the beginning of that century. We have explained in chapter 4 why mathematical analysis can have no firm basis without the theory of continuum; we have given there one of the simplest methods for evolving this theory. In the seventies of the last century, several such theories appeared simultaneously; they were all quite satisfactory, each had its own advantages and was equivalent to one another in the formal logical sense. Weierstrass, Dedekind and Cantor must be mentioned in connection with these theories.

The theory of real numbers cannot, of course, be regarded as part of mathematical analysis. It belongs to the theory of numbers and the theory of sets. However, numbers are the medium which originate and develop all concepts of mathematical analysis and therefore a thorough theoretical basis of mathematical analysis could not develop until the properties of this medium were studied to the end. Hence, only after the theory of continuum was finally created, mathematical analysis reached its present state.

V

It is, of course, obvious that apart from the revision and improvement of fundamentals the structure of mathematical analysis continued to develop as it does to the present day. During the first half of the XIX century the attention of scholars was strongly attracted by integral calculus. Thus integrability of elementary functions, new transcendental functions defined as primitives of elementary (in particular algebraic) functions or by means of integrals depending on parameters, the general theory of integrals of several dimensions

(multiple) and many other problems were subjected to strict studies.

However, from the beginning of the last century, the centre of gravity of scientific interests of analysts became progressively displaced towards higher sections of analysis, first of all the theory of differential equations—a problem which to this day occupies a central position among analytical problems. At the end of the last century, another problem was added to it, *viz.*, the theory of integral and integro-differential equations; this subject became immediately popular mainly because of its numerous practical applications. Finally, calculus of variations which developed systematically from the beginning of the XVIII century onwards was, in the course of the last few decades, regarded as a particular problem of a new and important scientific branch, *viz.*, functional analysis whose general development continues to attract more and more attention.

Hence, if we assume that mathematical analysis covers not only differential and integral calculus but also the whole set of the newly-created higher sections of the analytical science, then the horizons of this science widen greatly, and it is impossible to foresee the time when its problems may become exhausted; history tells us that before one circle of problems of mathematical analysis is completely solved many other problems arise which demand an immediate solution.

VI

From the XIX century onwards Russian mathematicians participated in the development of mathematical analysis—to begin with, individual scholars directed attention towards this field and later they were joined by powerful mathematical schools. The contribution by our scholars to this science during the XIX and the first half of the XX centuries is so significant that it must undoubtedly be considered separately, particularly since the contributions by Russian scholars in this field, apart from their great scientific value, are characterised by a special approach which differ considerably from that of foreign scholars.

It is well-known that our great geometry specialist N. I. Lobachevskij paid almost no attention to the problems of mathematical analysis; it is therefore more significant that he expressed views whose depth and insight beats the views held by specialists of this era. Thus, the modern definition of functional dependence which is usually connected with the name of Dirichlet and arose as a result of the victory of the real, concrete approach over the formalistic

approach (*cf.* § 4) was expressed several years earlier and formulated with great accuracy by Lobachevskij*; he clearly emphasizes that for functional dependence of y on x it is only necessary that a definite value of y should correspond to every value of x *regardless of the way in which this relationship is given*. And this is also the essence of Dirichlet's definition.

In this course we have met twice the name of the outstanding Russian mathematician M. V. Ostrogradskij. Apart from developing a remarkable method for integrating rational functions (§ 61) and the famous formula which expresses a triple integral in a three-dimensional region in terms of a double integral over the surface of this region**, Ostrogradskij also deduced several other results which are of fundamental importance in integral calculus. Thus he was also the first to prove the formulae for transformation of variables in multiple integrals and explain the part played in such transformations by the so-called "functional determinants" or "Jacobians" (this name is derived from the surname of the German mathematician Jacobi who studied the properties of these determinants after they were discovered by Ostrogradskij; the latter unfortunately failed to publish them***).

Ostrogradskij also made important investigations in analytical mechanics and calculus of variations. His works (apart from the fields mentioned above, he was also interested in ballistics, mechanics of heavenly bodies, theory of probability, theory of algebraic functions, etc.) are characterised by his deep interest in applied sciences and his attempts to place mathematical sciences on as wide a basis as possible, to express all problems in the most general form and then solve them strictly and accurately on their own merits.

Around the middle of the XIX century, works by the greatest Russian analyst P. L. Chebyshev began to appear. Chebyshev belonged to the school of mathematicians who succeeded in working with equal success and interest in many branches of mathematics. He investigated problems of integral calculus, approximation of functions in terms of polynomials with interpolations of various kinds, theory of numbers, theory of probability and theory of mechanisms;

* See B. V. Gnedenko, "Sketches from the History of Mathematics in Russia," Gostekhizdat, 1946, p. 96.

** M. V. Ostrogradskij solved this problem for space of any dimensions, *i.e.*, he established a general formula which replaced evaluation of an integral of multiplicity n by an integral of multiplicity $(n - 1)$.

*** In this text we have called them "Ostrogradskij's determinants".

and almost in every field he succeeded in developing new methods whose use continued for many years to be a model for his pupils and successors. His theories on approximation of functions in terms of polynomials continued to develop until at present they form a separate scientific branch—"constructive theory of functions". His works on the theory of probability have completely transformed the outlook of this science; he was the first to state its general problems and develop methods for their solution. In the theory of numbers Chebyshev was the first to develop the theory of distribution of simple numbers into a natural series which was left at a standstill for a long time; on the other hand, he laid foundations for solution of heterogeneous problems in the theory of diophantine approximations; he uncovered such a huge field of activity to succeeding generations that it has not yet been exhausted. Chebyshev's works on the theory of mechanisms have not lost their importance to this day, either in theory or in practice. In the field of mathematical analysis in which we are interested Chebyshev made a series of investigations in integrability of elementary functions which were difficult to integrate, and, in particular, he established the famous theorem on integrability of binomial differentials (§ 64). Chebyshev also published several important works on integration of rational functions, approximate evaluation of integrals, interpolation and the so-called "problem of moments".

The scientific approach in Chebyshev's works is mainly characterised by the tendency to solve practical problems. In his article "Drawing of Geographical Maps" Chebyshev wrote: "The relationship between theory and practice gives the most desirable results and is to the advantage of both practice and theory; science itself develops under its influence; it reveals new subjects for investigation or new sides of well-known subjects". The example which best illustrates Chebyshev's point of view is that he evolved the general theory of approximating functions in terms of polynomials as a result of solving one concrete problem in the theory of mechanisms. However, this example also describes another side of Chebyshev's scientific creativeness. Although strictly adhering to practical requirements, he never tried to solve one isolated case. On the contrary, he always tried to place such problems on the widest possible footing and deduce mathematical theories which would embrace the greatest number of similar problems. The history of development of mathematics shows that solution of practical problems is most useful for the development of the mathematical science as a whole.

Chebyshev formed the first large mathematical school in Russia which soon gained world-wide importance. Chebyshev's brilliant pupils (Zolotariev, Liapunov, Markov and others) partly continued his investigations, but they also tried to conquer new fields. In the history of mathematical analysis and in its physical applications, the remarkable work of A.M. Liapunov is of the greatest importance. Liapunov created a new trend in analysis which was mainly prompted by problems of mechanics and mathematical physics but which soon gained an independent mathematical meaning. The main objects of his investigations were equilibrium conditions of fluid bodies on the one hand (important in the study of heavenly bodies) and, on the other hand, the problems of stability and instability under equilibrium conditions and movement of mechanical systems. During this period (at the junction of the XIX and XX centuries), these problems were of universal interest; Liapunov also worked in conjunction with the famous French scientist Poincare who was also interested in these problems. It is interesting to note the different approach of these two scientists, since they are characteristic of the Russian school as a whole as compared to many West European schools. In solving physical problems, Poincare often did not permit strictly accurate assumptions and, realising this, he maintained that "you cannot require the same strictness in mechanics as in pure analysis". But Liapunov solved the same type of problems with absolute accuracy and said * "we must not use doubtful arguments no matter how soon they give us solution of the given problem, regardless of whether it is a problem of mechanics or physics, provided it is stated quite definitely from the analytical point of view. It thus becomes a problem of pure analysis and should be treated as such". It is therefore clear that because of this difference in treatment Liapunov's results have greater finality and are more fundamental in character than the achievements by the french scientist.

Liapunov was the first to prove a very important theorem on closed trigonometrical orthogonal systems (§ 83). In the applications of analysis he was the first to prove the so-called "central limit theorem" in the theory of probability which is still of great importance in this branch of mathematics. He carried out his proof with the help of a new original method whose general outlines were worked out much later and proved to be one of the most essential methods in the analytical theory of probability.

* *c.f.* Notes of the Academy of Sciences, Physico-mathematical section, 8th series 1905, vol. 17, No. 38, pp. 1-32.

Apart from Chebyshev's school, the scientific creativeness of S. V. Kovalevskaja must be mentioned. Among her works, two are of fundamental importance; one of them deals with the theory of differential equations and the other with the mechanical problem of movement of a solid body with a stationary point. Kovalevskaja worked mostly abroad, since as a woman she could not find suitable conditions for her work in Imperial Russia. Nevertheless, all her works have the typical characteristics of the Russian mathematical school. She reveals the same strict concern for her subject, the great interest in applied sciences and the same width and generality in defining the problems which she solved with an absolute accuracy of the technico-mathematical arguments involved; this is so characteristic of Chebyshev and of all his pupils that it gave the Russian mathematical school its peculiar monumental style which none of the West European schools could claim.

The next generation of Chebyshev's school worked partly in Soviet times. To this generation belong such leading representatives of mathematical analysis as V. A. Steklov and S. N. Bernstein who greatly contributed to the analytical treasury in the field of differential equations, constructive theory of functions and in many other branches as well as in applied sciences, *viz.*, mathematical physics and theory of probability.

The general growth of sciences in the USSR after the Great October Socialist Revolution raised the work on mathematical analysis both in quality and quantity to a higher level. Science, and therefore also mathematical sciences, now has many more workers than in pre-revolution days; on the other hand, the right and competent planning of research work, scientific establishments and scientific publications as well as thorough, planned and highly authoritative education of the forthcoming generation assures improved quality of scientific findings. The Soviet team of workers on mathematical analysis under the leadership of our academicians (S. N. Bernstein, M. V. Keldysh, N. M. Krylov, M. A. Lavrentiev, I. G. Petrovskij, V. I. Smirnov, S. L. Sobolev) already have to their credit a long string of first-rate achievements. Faithful to the famous traditions of Russian mathematics and inspired by the desire to give all their strength to their country and the Soviet people, they assuredly go forward towards new conquests.

INDEX

- Absolute convergence of an integral, 498.
- Absolute convergence of an integral series, 318.
- Acceleration, 136.
- Algebraic function, 14.
- Alternating series, 316.
- Approximate evaluation of integrals by the method of parabolas, 262.
- Approximate evaluation of integrals by the method of trapeziums, 257.
- Area of a curvilinear trapezium, 193.
- Area of a surface, 593.
- Average approximations, 388.

- Bernstein's polynomials, 374.
- Binomial differentials, 287.
- Bounded quantity, 24.
- Bounded set of numbers, 71.
- Bounds of a bounded set, 72.
- Boundary of a plane figure, 558.
- Boundary point of a plain figure, 558.

- Cauchy's form for the last term of Taylor's series, 160.
- Cauchy's theorem, 146.
- Centre of curvature, 458.
- Circle of curvature, 458.
- Circulation in a vector field, 651.
- Closed orthogonal system, 390.
- Closed region, 404, 558.
- Composite function, 85.
- Conditional convergence of series, 318.
- Conditional extrema, 483.
- Continuity of a function along a line, 83.
- Continuity of a function at a point, 80.
- Continuity of functions of several variables, 403.
- Continuity of sum of a series of functions, 335.
- Continuum, 63.
- Contour of a plane figure, 558.
- Contracting sequence of regions, 405.
- Contracting sequence of sections, 70.
- Convergence of Fourier's series, 380.
- Convergence of integrals with infinite limits, 491.
- Convergence of integrals of unbounded functions, 506.
- Convergence of sequences, 49.
- Convergence of series, 299.
- Criterion for convergence of an infinite product, 330.
- Criterion for existence of a limit, 76.
- Criterion of integrability, 208.
- Curvature of a plane curve, 453.
- Curvilinear integrals, 602.
- Curvilinear integrals in space, 622.

- Definite integrals, 201.
- Derivative, 108.
- Derivative in a given direction, 419.
- Derivatives inverse trigonometrical functions, 121.
- Derivative of a composite function, 117.
- Derivative of a logarithm, 116.
- Derivative of a power, 110.
- Derivative of a power function, 120.
- Derivative of a quotient, 113.
- Derivative of an algebraic sum, 111.
- Derivative of an exponential function, 120.
- Derivative of an implicit function, 426.
- Derivative of a product, 112.
- Derivative of an inverse function, 119.
- Derivative of higher orders, 136.
- Derivatives of trigonometrical functions, 114.
- Diameter of a region, 404.
- Differential of a function of one variable, 129.
- Differential of a function of two variables, 414.
- Differentials of higher orders, 139.
- Differentiability of functions of one variable, 130.
- Differentiability of functions of two variables, 418.

- Differentiability of functions of three variables, 419.
- Direction of convexity (concavity) of a curve, 451.
- Dirichlet's function, 9.
- Dirichlet-Liapunov theorem, 390.
- Discontinuity of functions, 81.
- Distance between two regions, 407.
- Distance of a point, from a region, 404.
- Divergence of a sequence, 49.
- Divergence of a series, 299.
- Divergence of vector field, 649.
- Double integral, 571.
- Elementary functions, 13.
- Equation of normal plane to a curve in space, 446.
- Equation of normal to a surface, 448.
- Equation of tangent normal to a plane curve, 443.
- Equation of tangent to a plane curve, 443.
- Equation of tangent to a plane curve in space, 446.
- Equation of a tangential plane to a surface, 448.
- Equivalent infinitely small quantities 40.
- Euler's integrals 541.
- Euler's substitutions, 286.
- Expansion of a rational fraction into simple fractions, 269.
- Exponential function 15.
- Exterior point of plane figure, 558.
- Extrema of function 167, 438.
- Finite coverage, 71.
- Flow of vector field across a surface, 649.
- Fourier coefficients, 377.
- Fourier's series, 380.
- Functions, 3, 7.
- Functional dependence, 5.
- Generalised trigonometrical series, 396.
- Gradient, 647.
- Graph of function, 11.
- Green's formula, 612.
- Homogeneous function, 427.
- Implicit functions, 442, 462.
- Increasing and decreasing functions, 164.
- Indefinite integral, 176.
- Independent variable, 5.
- Infinite products, 326.
- Infinitely large quantities, 26.
- Infinitesimal and infinitely large quantities of different orders, 39.
- Infinitesimal quantities, 18.
- Integral, 201.
- Integral logarithm, 515.
- Integral over a section of a curve, 240, 242.
- Integrals over a section of a curve part of a surface, 597.
- Integral sums, 202.
- Integral test for convergence of series, 502.
- Integral with infinite limits, 491.
- Integrals of unbounded functions, 504.
- Integrand, 203.
- Integrand expression, 203.
- Integration by parts, 183.
- Integration of an algebraical sum, 182.
- Integration of binomial differentials, 287.
- Integration of differentials containing exponential functions, 294.
- Integration of simple fractions, 274.
- Integration of trigonometrical differentials, 289.
- Integrability of functions of two variables, 572.
- Interior point of plane figure, 557.
- Interval of integration, 203.
- Invariance of first differential, 134.
- Inverse function, 91.
- Inverse trigonometrical functions, 16.
- Irrational numbers, 60.
- Irregular rational fraction, 266.
- L'Hospital's law, 148.
- Lagrange's form for the last term of Taylor's series, 160.
- Lagrange's theorem, 144.
- Last term in Taylor's formula, 158.
- Law of motion, 102.
- Leibnitz's formula, 139.
- Leibnitz's theorem on alternating series, 316.
- Length of arc of a plane curve, 231, 234.

- Length of arc of a plane curve in space, 241.
- Level surfaces, 647.
- Limit of a variable quantity, 53.
- Limit of a variable sequence, 49.
- Limits of integration, 203.
- Local extrema, 167, 438.
- Local property, 82.
- Logarithmic function, 16.

- Maclaurin's formula, 157.
- Maclaurin's series, 362.
- Mathematical definition of a process... 45.
- Maximum of function, 168.
- Measure of plane figures, 561.
- Measurability of a plane figure, 561.
- Minimum of a function, 168.

- Natural logarithms, 117.
- Number plane, 401.

- One-sided continuity, 83.
- One-sided limit of a function, 50.
- Open region, 404, 558.
- Operations with infinitely small quantities, 23.
- Orthogonality of functions, 378.
- Ostrogradskij's determinant, 472.
- Ostrogradskij's formula, 637.
- Ostrogradskij's method for integrating rational fractions, 277.

- Partial derivatives, 410.
- Partial sums of higher orders, 429.
- Partial sums of series, 299.
- Plane figures, 557.
- Point of continuity and discontinuity of a function, 83.
- Point of convergence, divergence of a series, 334.
- Point of inflexion, 453.
- Polynomials, 13.
- Potential of vector field, 651.
- Power function, 14, 351.
- Primitive, 175, 176.
- Principal Linear part of the increment of functions, 130.
- Principle of comparison of series, 307.
- Properties of integrals, 213, 223.
- Properties of Ostrogradskij's determinants, 475.

- Radius of convergence of a power series, 351.
- Radius of curvature, 458.
- Rational fraction, 266.
- Rational function, 14.
- Rational integral function, 13.
- Rational numbers, 60.
- Rationalisation of an integrand, 282.
- Real numbers, 63.
- Region of convergence, divergence of series, 33.
- Region of convergence of power series, 351, 353.
- Region of definition of a function, 6.
- Region of integration, 571.
- Regular rational fraction, 265.
- Relationship between an integral and a primitive, 213.
- Remainder of series, 300.
- Replacement of variables in an integral, 187.
- Replacement of variables in a double integral, 584.
- Rolle's theorem, 144.
- Rotor of a vector field, 652.

- Scalar field, 647.
- Sequences of numbers, 48, 49.
- Series of functions, 333.
- Series of polynomials, 369.
- Smooth curve, 240, 242.
- Stationary point, 169, 438.
- Stirling's formula, 548.
- Stoke's formula, 642.
- Straightening curves, 237, 242.
- Substitution method for integrating functions, 187.
- Sum of an infinite series, 298.
- Surface integrals, 626.
- Surface of body of rotation, 251.

- Tangential circle, 458.
- Taylor's formula, 154.
- Taylor's formula for function of two variables, 433.
- Taylor's series, 364.

- Term-by-term addition and subtraction of series, 304.
- Term-by-term differentiation of series, 344.
- Term-by-term integration of series, 344.
- Test for the convergence of Cauchy's series, 308.
- Test for the convergence of D'Alembert's series, 309.
- Test for the convergence of Dirichlet's series, 319.
- Test for the convergence of Raabe Series, 313.
- Theorem on mean values for integrals, 228.
- Theorem on mean values for double integrals, 573.
- Transcendental function, 15.
- Trigonometrical function, 16.
- Trigonometrical polynomial, 387.
- Trigonometrical series, 377.
- Turbulent field, 652.
- Two-dimensional continuum, 403.
- Two-sided limit of a function, 55.
- Uniform continuity of power series, 351.
- Uniform continuity series of functions, 334.
- Uniform continuity of functions of one variable, 92.
- Uniform continuity of functions of two variables, 409.
- Uniform convergence of an integral, 526.
- Uniform convergence of a sequence, 341.
- Variable quantity, 2.
- Variation of a function in an interval, 208.
- Vector field, 647.
- Vector of turbulence, 651.
- Velocity of uniform movement, 101.
- Volume of a body of rotation, 250.
- Weierstrass's theorem, 372.

